UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Empirical Identification and Longitudinal Characterization of Mild Cognitive Impairment Subtypes using Latent Mixture Modeling

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Clinical Psychology

by

Joel Stephen Eppig

Committee in charge:

University of California San Diego

Professor Mark W. Bondi, Chair
Professor Lisa Delano-Wood
Professor Emily C. Edmonds

San Diego State University

Professor Paul E. Gilbert
Professor Scott C. Roesch

2019

The Dissertation of Joel Stephen Eppig is approved, and it is acceptable

in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

San Diego State University

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to thank my wonderful graduate advisor, Dr. Mark Bondi, who has provided invaluable mentorship, encouragement, and opportunities for me throughout my graduate career. It has been a privilege to learn from and train with you, and this dissertation would not have been possible without your continued support and guidance of my professional and scientific ambitions. My sincere gratitude also goes to the other members of my dissertation committee, Drs. Lisa Delano-Wood, Emily Edmonds, Paul Gilbert, and Scott Roesch, whose helpful guidance and feedback has allowed me to see this project through to completion. I would like to express my appreciation to Dr. Emily Edmonds, Dr. Kelsey Thomas, Mark Sanderson M.S., and Laura Campbell, B.S for their integral roles in assisting with the compilation and management of ADNI data as well as scoring and standardization of measures. A special thanks to Dr. Michael Taylor, who generously provided his time and expertise on the development and psychometrics of advanced normative methodologies.

I would also like to express my appreciation to the numerous individuals who have helped me along my journey in myriad ways during graduate school. First and foremost, a special thank you to my wife, Bahiah, for her unconditional love and partnership; I am endlessly grateful for her unwavering patience, support, and encouragement of my pursuits, both during my graduate career and the dissertation process. I would also like to thank my parents for their support of my professional and personal ambitions, and who taught me to always be the best person I can be. Finally, I'd like to express my gratitude to all of my incredible JDP classmates, colleagues, and labmates, who have helped me grow and be a better researcher, clinician, and person.

Study 1, in part, is a reprint of the material as it appears in the *Journal of the International Neuropsychological Society, (2017), 23*(7), 564-576. Eppig, Joel S.; Edmonds, Emily C.; Campbell, Laura; Sanderson-Cimino, Mark; Delano-Wood, Lisa; and Bondi, Mark W. for the Alzheimer's Disease Neuroimaging Initiative, Cambridge University Press. The dissertation author was the primary investigator and author of this paper.

VITA

2013 – 2019	Doctor of Philosophy in Clinical Psychology, San Diego State University / University of California San Diego Joint Doctoral Program in Clinical Psychology
2018 – 2019	Internship in Clinical Psychology, VA Puget Sound Healthcare System
2016	Master of Science in Clinical Psychology, San Diego State University
2009	Bachelor of Arts in Psychology, Minors in Cognitive Neuroscience & Music Temple University

PUBLICATIONS

1. Thomas, K. R., Bangen, K. J., Weigand, A. J., Edmonds, E. C., Sundermann, E., Wong, C. G., **Eppig, J.,** Werhane, M. L., Delano-Wood, L., & Bondi, M. W. (in press). Type II diabetes interacts with Alzheimer's disease risk factors to predict functional decline. *Alzheimer's Disease & Associated Disorders.*

2. Edmonds, E. C., McDonald, C. R., Marshall, A., Thomas, K. R., **Eppig, J.,** Weigand, A. J., Delano-Wood, L., Galasko, D. R., Salmon, D. P., & Bondi, M. W. (in press). Early vs. late MCI: Improved MCI staging using a neuropsychological approach. *Alzheimer's & Dementia.* doi: 10.1016/j.jalz.2018.12.009

3. Bangen, K. J., Weigand, A. J., Thomas, K. R., Delano-Wood, L., Clark, L. R., **Eppig, J.,** Werhane, M. L., Edmonds, E. C., & Bondi, M. W. (in press). Cognitive dispersion is a sensitive marker for early neurodegeneration changes and functional decline in nondemented older adults. *Neuropsychology.* doi.org/10.1037/neu0000532

4. Thomas, K. R., **Eppig, J.,** Weigand, A. J., Edmonds, E. C., Wong, C. G., Jak, A. J., Delano-Wood, L., Galasko, D. R., Salmon, D. P., Edland, S. D., & Bondi, M. W. (2019) Artificially low MCI-to-normal reversion rate in ADNI. *Alzheimer's & Dementia, 15*(4), 561-569. doi: 10.1016/j.jalz.2018.10.008

5. Edmonds, E. C., Weigand, A. J., Thomas, K. R., **Eppig, J.**, Delano-Wood, L., Galasko, D. R., Salmon, D. P., & Bondi, M. W. (2018). Increasing inaccuracy of self-reported subjective cognitive complaints over 24 months in empirically-derived subtypes of mild cognitive impairment. *Journal of the International Neuropsychological Society, 24*(8)*,* 842-853. doi:10.1017/S1355617718000486

6. Wong, C. G., Thomas, K. R., Edmonds, E. C., Weigand, A. J., Bangen, K. J., **Eppig, J.,** Jak, A. J., Devine, S. A., Delano-Wood, L., Libon, D. J., Edland, S. D., Au, R., & Bondi, M. W. (2018). Neuropsychological criteria for mild cognitive impairment in the Framingham Heart Study's old-old. *Dementia & Geriatric Cognitive Disorders, 46(5-6)*, 253-365. doi:10.1159/000493541

7. Thomas, K. R., Edmonds, E. C., **Eppig, J.,** Salmon, D. P., & Bondi, M. W. (2018). Using neuropsychological process scores to identify subtle cognitive decline and predict progression to MCI. *Journal of Alzheimer's Disease, 64*(1), 195-204. doi:10.3233/JAD-180229

8. Thomas, K. R., **Eppig, J.,** Edmonds, E. C., Jacobs, D. M., Libon, D. J., Salmon, D. P., & Bondi, M. W. (2018). Word-list intrusion errors predict progression to mild cognitive impairment. *Neuropsychology, 32*(2), 235-245. doi:10.1037/neu0000413

9. Roll, E., Giovannetti, T., Libon, D. J., & **Eppig, J.** (2017). Everyday task knowledge and everyday function in dementia. *Journal of Neuropsychology. 13*(1), 96-120. doi:10.1111/jnp.12135

10. **Eppig, J.,** Edmonds, E. C., Campbell, L. M., Sanderson-Cimino, M., Delano-Wood, L., & Bondi, M. W. (2017). Statistically-derived subtypes and associations with cerebrospinal fluid and genetic biomarkers in mild cognitive impairment: A latent profile analysis. *Journal of the International Neuropsychological Society, 23*(7), 564-576. doi:10.1017/S135561771700039X

11. Edmonds, E. C., **Eppig, J.,** Bondi, M. W., Leyden, K. M., Goodwin, B., Delano-Wood, L., & McDonald, C. R. (2016). Heterogeneous cortical atrophy patterns in MCI not captured by conventional diagnostic criteria. *Neurology*, *87*(20), 2108-2116. doi:10.1212/WNL.0000000000003326

12. Libon, D. J., **Eppig, J.,** Drabick, D., Erwin, K., Durmala, L., Bloch, A., Donohue, J., Swenson. R., Alexander, G., Aradillas, E., Grothusen, J., Peterlin, B. L., & Schwartzman, R. J. (2016). Neuropsychological and related clinical features associated with Complex Regional Pain syndrome: A latent class analysis. *British Journal of Medicine & Medical Research, 18*(12), 1-12. doi:10.9734/BJMMR/2016/29713

13. Ahmed, S., Brennan, L., **Eppig, J.,** Price, C. C., Lamar, M., Delano-Wood, L., Bangen, K. J., Edmonds, E. C., Clark, L., Nation, D. A., & Jak, A. (2016). Visuoconstructional impairment in subtypes of mild cognitive impairment. *Applied Neuropsychology: Adult*, *23*(1), 43-52. doi:10.1080/23279095.2014.1003067

14. Libon, D. J., Drabick, D., Giovannetti, T., Price, C. C., Bondi, M. W., **Eppig, J.,** Nieves, C., Wicas, G., Lamar, M., Delano-Wood, L., Nation, D. A., Brennan, L., Au, R., & Swenson, R. (2014). Neuropsychological syndromes associated with Alzheimer's/ vascular spectrum dementia: A latent class analysis. *Journal of Alzheimer's Disease*, *42*(3), 999-1014. doi:10.3233/JAD-132147

15. Ayaz, H., Shewokis, P. A., Scull, L., Libon, D. J., Feldman, S., **Eppig, J.,** Onaral, B., & Heiman-Patterson, T. (2014). Assessment of prefrontal cortex activity in Amyotrophic Lateral Sclerosis patients with functional near infrared spectroscopy. *Journal of Neuroscience and Neuroengineering*, *3*(1), 41-51. doi:10.1166/jnsne.2014.1095

16. Seidel, G., Giovannetti, T., Price, C. C., Tanner, J., Mitchell, S., **Eppig, J.,** & Libon, D. (2013). Neuroimaging correlates of everyday action in dementia. *Journal of Clinical and Experimental Neuropsychology, 35*(9)*,* 993-1005*.* doi:10.1080/13803395.2013.844773

17. Giovannetti, T., Britnell, P., Brennan, L., Siderowf, A., Grossman, M., Libon, D. J., Bettcher, B. M., Rouzard, F., **Eppig, J.,** & Seidel, G. (2012). Everyday action impairment in Parkinson's disease dementia. *Journal of the International Neuropsychological Society, 18*(5), 787-798. doi:10.1017/S135561771200046X

18. **Eppig, J.**, Wambach, D. M., Nieves, C., Price, C. C., Lamar, M., Delano-Wood, L., Giovannetti, T., Bettcher, B. M., Penney, D.L., Swenson, R., Lippa, C., Kabasakalian, A., Bondi, M. W., & Libon, D. J. (2012). Dysexecutive functioning in mild cognitive impairment: Derailment in temporal gradients. *Journal of the International Neuropsychological Society, 18*(1)*,* 20-28. doi:10.1017/S1355617711001238

19. Libon, D. J., Bondi, M. W., Price, C. C., Lamar. M., **Eppig, J.,** Wambach, D. M., Nieves, C., Delano Wood, L., Giovannetti, T., Lippa, C., Kabasakalian, A., Cosentino, S., Swenson, R., & Penney, D. L. (2011). Verbal serial list learning in mild cognitive impairment: A profile analysis of

interference, forgetting, and errors. *Journal of the International Neuropsychological Society, 17*(5)**,** 905-914. doi:10.1017/S1355617711000944

20. Bettcher, B. M., Giovannetti, T., Libon, D. J., Wambach, D. M., **Eppig, J.**, & Klobusicky, J. (2011). To err is human, to monitor divine: Environmental adaptation reduces errors but do not improve monitoring. *Journal of Clinical and Experimental Neuropsychology, 33*(10), 1049-1058. doi:10.1080/13803395.2011.595395

21. Bettcher, B. M., Giovannetti, T., Libon, D. J., Wambach, D. M., **Eppig, J.**, & Klobusicky, J. (2011). Improving everyday error detection one picture at a time: A performance-based study of everyday task training. *Neuropsychology, 25*(6), 771-783. doi:10.1037/a0024107

22. Libon, D. J., Schwartzman, R. J., **Eppig, J.**, Wambach, D. M., Brahin, E, Peterlin B. L., Alexander, G., & Kalanuria, A. (2010). Neuropsychological deficits associated with Complex Regional Pain Syndrome. *Journal of the International Neuropsychological Society, 16*(3), 566-573. doi:10.1017/S1355617710000214

23. Libon, D. J., Xie, S. X., **Eppig, J.**, Wicas, G., Lippa, C., Bettcher, B. M., Price, C. C., Giovannetti, T., Lamar, M., Swenson, R., & Wambach, D. M. (2010). The heterogeneity of mild cognitive impairment: A neuropsychological analysis. *Journal of the International Neuropsychological Society, 16*(1), 84-93. doi:10.1017/S1355617709990993

BOOK CHAPTERS

1. **Eppig, J.,** Werhane, M., Edmonds, E. C., Delano-Wood, L., Bangen, K. J., Jak, A. J., Thomas, K. R., Wong, C. G., Weigand, A. J., & Bondi, M. W. (in press). Neuropsychological diagnostic criteria for mild cognitive impairment associated with Alzheimer's disease. In D.J. Libon, M. Lamar, K.M. Heilman, & R. Swenson (Eds.), *Vascular Disease, Alzheimer's Disease, and Mild Cognitive Impairment: Advancing an Integrated Approach.* New York, NY: Oxford University Press.

2. Libon, D. J., Lamar, M., Price, C. C., Garrett, K. D., Giovannetti, T., Delano-Wood, L., Penney, D. L., **Eppig, J.**, Nieves, C., Bondi, M. W., & Swenson, R. (2013). The dysexecutive syndrome associated with Alzheimer/vascular spectrum dementia syndromes and mild cognitive impairment: An analysis of process and errors. In L. Ashendorf, R. Swenson, & D.J. Libon (Eds.), *The Boston Process Approach to Neuropsychological Assessment: A Practitioner's Guide* (pp. 134-147). New York, NY: Oxford University Press.

ABSTRACT OF THE DISSERTATION

Empirical Identification and Longitudinal Characterization of Mild Cognitive Impairment Subtypes using Latent Mixture Modeling

by

Joel Stephen Eppig

Doctor of Philosophy in Clinical Psychology

University of California San Diego, 2019
San Diego State University, 2019

Professor Mark W. Bondi, Chair

Rationale: Research in conventional mild cognitive impairment (MCI), a prodromal stage between normal aging and Alzheimer's dementia (AD), has demonstrated neuropsychological heterogeneity using clustering techniques. The current dissertation aimed to 1) empirically establish baseline neuropsychological MCI subtypes; 2) explore longitudinal characterization of empirical subtypes using rigorous norms; and 3) examine the probability of transition between subtypes over time.

Design: Study 1 included 806 MCI participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Unique neuropsychological MCI subtypes and their associations with AD markers were

investigated using latent profile analysis (LPA). Study 2 included 825 ADNI participants with baseline MCI that had follow-up at 12-months (n=751) and 24-months (n=639). Demographically-corrected T-scores were derived from the performance of 284 “robust” normal control participants assessed at baseline, 12-, and 24-months. Serial LPAs established neuropsychological subtypes for the MCI participants at each time point. Study 3 employed latent transition analysis to evaluate the likelihood of subtype change over time, as well as the influence of AD-risk factor covariates on transition probabilities.

Results: Study 1 produced 3-classes: mixed impairment, amnestic impairment, and cognitively normal neuropsychological subtypes. Amnestic and mixed classes had higher positivities on markers of AD than the cognitively normal class. In Study 2, 4-neuropsychological classes were separately established at baseline, 12-, and 24-months: multi-domain impairment ([MLT]), amnestic impairment (AMN), dysexecutive/below average cognition (DYS/BA), and average cognition (AVG) classes. The MLT and AMN subtypes declined over time on the majority of measures, while the AVG subtype had stable neurpsychological performance. The DYS/BA subtype demonstrated stable memory performance and improvement on language and attention/executive measures. Study 3 indicated a high probability (>86%) for participants of all subtypes to remain in their class over time. Covariates that modestly increased the likelihood of transition between classes included worse functional ability and AD-biomarker positivity.

Conclusions: This dissertation research used latent mixture models to establish analogous longitudinal neuropsychological profiles in conventionally diagnosed MCI. Results suggest that individuals are most likely to remain within their subtype across two years, including cognitively normal “false-positives.” Future studies should examine empirical MCI subtypes with the use of actuarial methods that may improve diagnostic accuracy.

I. Introduction

Importance and Background of Mild Cognitive Impairment

Alzheimer's disease (AD) is a form of neurodegenerative dementia characterized by progressive impairment in cognitive and functional abilities (McKhann et al., 2011). Currently, estimates suggest there are 5.3 million older adults with AD dementia in the United States, and – due to an aging baby boomer population – that number is projected to increase to 13.8 million by 2050 (Alzheimer's Association, 2017). In addition to the devastating personal impact AD has on individuals and families, AD also imposes a significant societal burden through high healthcare costs and caregiving needs. However, the vast majority of pharmaceutical trials over the past 30 years have failed to produce any meaningful treatment that significantly impacts the progression of AD (Cummings, Morstorf & Zhong, 2014). Moreover, scientists increasingly recognize that the best chance of treating AD may ultimately lie in prevention, such that neurodegenerative processes are targeted well before the presence of frank dementia (Sperling et al., 2014).

Mild cognitive impairment (MCI) is conceptualized as an intermediate stage between normal aging and AD, such that individuals are at increased risk to develop dementia (Petersen et al., 2001). The term mild cognitive impairment entered the literature in the 1980's, as researchers began to identify features of an AD prodrome (Reisberg et al., 2008). However, the conventional definition of MCI originates with the criteria proposed by Petersen et al. (2001). Initially, Petersen's operationalization of MCI focused solely on the presence of memory deficits and complaints within the context of intact activities of daily living (Petersen et al., 1999; Petersen, et al., 2001). However, subsequent revisions expanded the definition to include distinctions with respect to "amnesic" versus "non-amnesic" and "single" versus "multi-domain" cognitive dysfunction (Petersen, 2004; Winblad et al., 2004). Currently, The National Institute on Aging-Alzheimer's Association (NIA-AA) core clinical criteria for MCI due to AD include 1) Concern in changes over cognition; 2) Impairment in one or more cognitive domains; 3) Preservation of independence in functional abilities; and 4) Not demented (Albert et al., 2011).

Neuropsychological Heterogeneity in Conventional MCI

Despite recent improvements to the definition of MCI, lingering concerns remain over how to operationalize specific criteria. Conventionally, the evaluation of cognitive impairment in MCI relies on subjective rating scales, brief cognitive screens, and/or a single impaired test score (often a delayed trial of story recall) rather than comprehensive neuropsychological assessment (Bondi et al., 2014). This diagnostic approach is evident among clinical trials targeting MCI (Petersen & Morris, 2005) as well as large-scale studies such as the Alzheimer's Disease Neuroimaging Initiative ([ADNI]; Weiner et al., 2013), despite research demonstrating the shortcomings of these methods (Bondi et al., 2014; Jak et al., 2009).

In an effort to address this issue, recent research has attempted to empirically classify distinct neuropsychological subtypes among individuals with conventional MCI. Early work employed cluster analytic methods to reveal heterogeneous cognitive profiles (Bondi et al., 2014; Delano-Wood et al., 2009; Clark, Delano-Wood, et al., 2013; Edmonds et al., 2015; Edmonds et al., 2016; Libon el al., 2010), challenging the empirical validity of the conventional diagnostic approach. These studies have consistently demonstrated evidence of three to four MCI subtypes depending on the cognitive domains examined, tests included, and diagnostic criteria. More recently, researchers have applied latent mixture models, such as Latent Profile Analysis (LPA) and Latent Class Analysis (LCA), to classify cognitive subtypes in MCI (Hanfelt et al., 2011; Köhler et al., 2013; McGuinness et al., 2015), which offer a model-drive approach and several statistical advantages over cluster analysis (Berlin, Williams, & Parra, 2014; Magidson & Vermunt, 2002; Muthén, 2004; Roesch, Villodas, & Villodas, 2010). These studies have identified three to seven neuropsychological classes, although some studies also included non-cognitive measures as latent indicators or normal older adults with subjective cognitive complaints as participants (Hanfelt et al., 2011; Köhler et al., 2013). A summary of the results of cluster analytic and latent mixture models approaches to empirical MCI classification are provided in Table 1.

Table 1: Results of cluster analytic and mixture model approaches in identifying empirical subtypes in conventional mild cognitive impairment.

Study	Statistical Model	Neuropsychological Domains	Sample Size	Number of Subgroups	MCI Subtypes (% of Sample)
Clark et al. (2013)	Ward's Cluster Analysis	Attention; Episodic Memory; Executive Functions; Language; Visuospatial Ability	n= 134	4	Cognitively Normal (47.8%) Memory/ Language (29.1%) Mixed/ Intact Attention (23.1%)
Delano-Wood et al. (2009)	Ward's Cluster Analysis	Constructional Praxis; Episodic Memory; Executive Functions; Language; Processing Speed	n= 70	3	Amnestic (48.6%) Memory/ Language (25.7%) Executive/ Processing Speed (25.7%)
Edmonds et al. (2015)	Ward's Cluster Analysis	Attention & Executive Functions; Episodic Memory; Language	n= 825	4	Amnestic (34.9%) Cognitively Normal (34.2%) Memory/ Language (18.5%) Executive/ Mixed (12.4%)
Hanfelt et al. (2011)	Latent Profile Analysis	Attention; Depression; Episodic Memory; Executive Functions; Functional Ability; Global Cognition; Processing Speed; Psychiatric Symptoms	n= 1,655	7	Executive/ Language (18%) Amnestic (16%) Amnestic + Functional/ Psych (16%) Functional + Psych (15%) Mixed (12%) Mixed + Functional/ Psych (12%) Cognitively Normal (12%)
Köhler et al., (2013)	Latent Profile Analysis	Executive Functions; Episodic Memory; Language; Processing Speed	n= 635	5	Cognitively Normal – Average (38%) Memory Retrieval (36%) Cognitively Normal – Above Average (15%) Executive/Processing Speed (5%) Amnestic (4%)
Libon et al. (2010)	K-means Cluster Analysis	Episodic Memory; Executive Functions; Language	n= 77	3	Amnestic (48.6%) Executive/ Language (27.3%) Mixed (25.7%)
McGuinness et al. (2015)	Latent Profile Analysis	Attention & Executive Functions; Episodic Memory; Language; Visuospatial Ability	n= 139	3	Mixed (40.3%) Amnestic (36.7%) Cognitively Normal (23.0%)

Note: Köhler et al. (2013) includes non-demented participants with "subjective or objective cognitive impairment on neuropsychological assessment." Hanfelt et al. (2011) used neuropsychological measures as well as questionnaires of functional abilities, neuropsychiatric symptoms, and depression as latent model indicators.

Within the ADNI corpus, Edmonds et al. (2015) was the first study to empirically classify conventionally diagnosed MCI subjects (n=825) using cluster analysis. Results produced four unique cognitive phenotypes: an amnestic MCI group (34.9%), a dysnomic MCI group (18.5%), a dysexecutive MCI group (12.5%) and a large fourth cluster (34.2%) characterized by intact neuropsychological performance *despite* their MCI diagnosis based on ADNI criteria. The "cluster-derived normal" group performed within normal limits on all neuropsychological cluster measures despite subjective complaints and impaired scores on the Wechsler Memory Scale-Revised (WMS-R) Logical Memory-II Story A and the Clinical Dementia Rating (CDR) scale that led to their ADNI MCI diagnosis. The notion that

individuals in this group were assigned a diagnosis of MCI in error was further supported by normal cerebrospinal fluid (CSF) Alzheimer's disease (AD) biomarker profiles and low rates of progression to AD and high rates of reversion to "cognitively normal" diagnoses (Bondi et al., 2014; Edmonds et al., 2015).

Inclusion of Visuoconstructional Ability

Despite the significant findings from Edmonds et al. (2015), there remain several unexplored clinical and statistical considerations that demonstrate the need for further study of the ADNI MCI cohort. First, the authors only examined the neuropsychological domains of attention/executive functions, language, and episodic memory, omitting any form of visuospatial skills such as visuoconstructional ability (which integrates visuospatial, organizational, and motor skills). In research studies, standardization and scoring of visuoconstructional measures can be time-consuming and prone to low inter-rater reliability, which may contribute to variable psychometric properties. However, in clinical practice, visuoconstructional ability is routinely assessed in the neuropsychological evaluation of older adults (Grossi & Trojano, 2001; Lezak, 2012). Additionally, significant visuospatial/constructional deficits are quite common among neurodegenerative disorders and dementia syndromes (Freedman & Dexter, 1991; Geldmacher, 2003) and represent an important component in neuropsychological protocols.

For example, Nielson, Cummings & Cotman (1996) demonstrated in autopsy-confirmed AD subjects a significant correlation between impaired visuoconstructional ability and hyperphosphorylated tau in occipital cortex. Moreover, visuoconstructional ability was not correlated with hyperphosphorylated tau in other brain regions, and language and memory functions were unrelated to hyperphosphorylated tau in occipital cortex. Prominent, differential visuospatial impairment is also a core diagnostic criterion of posterior cortical atrophy, a syndrome often attributable to AD pathology (Crutch et al., 2012; Crutch et al., 2013), and represents a key neuropsychological feature of Lewy body dementia (Ferman et al., 2006; Hamilton et al., 2008; Johnson, Morris & Galvin, 2005; Kao, et al., 2009; McKeith et al., 1996). Furthermore, individuals with non-amnestic MCI who progress to pathologically-confirmed Lewy body dementia have been shown to initially present with visuospatial/constructional as well as attentional impairments (Ferman et al., 2013; Molano et al., 2010). Visuospatial dysfunction has also been reported in multi-domain amnestic MCI (Mapstone, Steffenella & Duffy, 2003).

Importantly, a cluster analysis of amnestic and non-amnestic MCI subjects by Clark et al. (2013) revealed four unique subtypes, with three demonstrating visuoconstructional impairment: a single-domain visuoconstructional MCI subgroup (23.8%); an MCI subgroup with predominant executive and visuoconstructional dysfunction (16.3%); and a multi-domain MCI subgroup with mixed episodic memory, executive function, language and visuoconstructional impairment (17.5%). The fourth MCI subgroup was characterized by single-domain amnestic impairment only (42.5%), a consistent finding among all previous MCI neuropsychological classification studies (Delano-Wood et al., 2009; Edmonds et al., 2015; Libon et al., 2010). However, results in the visuospatial domain lack replication due to the exclusion of any representative assessment in the MCI classification literature. Thus, the contribution of visuoconstructional measures available in ADNI has potentially been overlooked by past studies identifying neuropsychological MCI subtypes (Bondi et al., 2014; Edmonds et al., 2015) and warrants additional study.

Benefits of Latent Mixture Models Over Cluster Analysis

A limitation of Edmonds et al. (2015) involves the use of traditional cluster analysis to identify subgroups. Newer latent mixture models, such as latent profile analysis (LPA), offer several statistical advantages over traditional cluster analysis given its model-driven classification approach. For example, while cluster analysis assigns each individual to subgroups in binary fashion, LPA utilizes maximum likelihood estimation to generate posterior probabilities and model the classification uncertainty of each individual in each latent class (Berlin, Williams, & Parra, 2014; Magidson & Vermunt, 2002; Muthén, 2004). These posterior probabilities are used to account for measurement error, consequently decreasing estimation bias and improving the accuracy of standard errors in analyses (Asparouhov & Muthén, 2015; Bray, Lanza, & Tan, 2015; Clark & Muthén, 2009; Magidson & Vermunt, 2002). LPA also produces information criterion and likelihood fit indices to guide determination of the number of optimal classes (Berlin, Williams & Parra, 2014; Muthén, 2004). This statistical comparison of nested models inherently increases objectivity and minimizes the arbitrary nature of subgroup selection in cluster analysis (Magidson & Vermunt, 2002). Other benefits of LPA include the ability to handle missing data points in analyses (Roesch, Villodas, & Villodas, 2010), accommodation of multiple data types such as categorical

and continuous variables (Magidson & Vermunt, 2002), incorporation of predictor variables and distal outcomes in the model (Magidson & Vermunt, 2002; Muthén, 2004), and model verification with independent samples (Shao, Liang, Yuan, & Bian, 2014).

Longitudinal Exploration of Empirical MCI Subtypes

Although a few studies have effectively used LPA to underscore the considerable heterogeneity in conventional MCI, almost all research has focused on classifying MCI at a single, initial evaluation (Hanfelt et al., 2011; Köhler et al., 2013; McGuinness et al., 2015). Thus, the longitudinal consistency and stability of MCI classes established at baseline remains unclear, as is the possibility of changes in subgroup size or interpretation. Currently, only one study has attempted to longitudinally categorize MCI subtypes. Peraita et al. (2015) performed serial LCAs across three years in a Spanish sample of community dwelling, healthy older adults without a prior diagnosis of dementia or MCI. Participants were administered a comprehensive neuropsychological battery at three time points and diagnosed via conventional criteria as amnestic MCI, non-amnestic MCI, or multi-domain MCI based on performance. Scores on 17 neuropsychological variables were then dichotomized as "adequate" or "low/poor" and used as latent class indicators in three separate, serial LCAs. The authors demonstrated a stable 4-class solution at each time point, with a healthy non-impairment class, amnesic class, non-amnesic class, and multi-domain impairment class present at all three years. However, the proportion of individuals in the healthy class decreased with time (baseline: 46.6%, year 1: 36.5%, year 2: 33.3%) while the amnesic (12.8%, 23.9%, 22.7%, respectively) and non-amnesic (21.7%, 21.2%, 24.5%, respectively) classes grew in size. The multi-domain impairment class remained largely consistent across all years (18.9%, 18.4%, 19.6%, respectively).

Despite these important contributions exploring neuropsychological class stability in elderly adults over time, there are also several limitations in the Peraita et al. (2015) study. Specifically, the authors categorized neuropsychological performance as "adequate" or "low/poor" rather than using standardized scores, arbitrarily setting the cut-off at the 40th percentile of uncorrected raw score performance. Two immediate problems are apparent with this methodology: 1) the lack of demographic adjustment, and 2) the use of a 40th percentile cut-score, which likely increases sensitivity at the considerable cost of

unsatisfactory specificity. Additionally, these issues are compounded by the fact that the authors did not utilize a separate normative control group, instead classifying neuropsychological performance based on the same sample subsequently used in the LCA.

In addition to a dearth of research characterizing neuropsychological performance of empirical MCI subtypes over time, no studies have attempted to investigate how empirical MCI classes relate to each other over time. Latent transition analysis (LTA), an extension of LPA, is a type of mixture model ideally suited to statistically address this issue. LTA models longitudinal change in class membership by auto-regressing one categorical latent variable onto another (Graham, Collins, Wugalter, Chung, & Hansen, 1991; Nylund, 2007), and can therefore examine transitions between MCI classes over time. For instance, Gu (2016) applied LTA to a mixed subsample of normal, MCI, and AD participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI), using two AD-associated neuroimaging markers (i.e., hippocampal volume and fluorodeoxyglucose-positron emission tomography [FDG-PET]) as indicators at baseline and year 2. The authors reported six classes at both time points and described several pathological possibilities of transition involving a reduction in either hippocampal volume or FDG-PET. However, research has yet to methodically investigate the longitudinal stability of empirical neuropsychological profiles in MCI and subsequently examine change in subtype classification over time.

Study 1, in part, is a reprint of the material as it appears in the *Journal of the International Neuropsychological Society, (2017), 23*(7), 564-576. Eppig, Joel S.; Edmonds, Emily C.; Campbell, Laura; Sanderson-Cimino, Mark; Delano-Wood, Lisa; and Bondi, Mark W. for the Alzheimer's Disease Neuroimaging Initiative, Cambridge University Press. The dissertation author was the primary investigator and author of this paper.

II. Purpose and Specific Aims

Goals of the Dissertation Research

The over-arching objective of this dissertation is to empirically identify and longitudinally characterize MCI subtypes in ADNI using latent mixture modeling. Multiple steps are required to achieve this objective, and this dissertation research is accordingly divided into three primary goals, each corresponding to one of three studies. As major steps in a larger project, these goals are interrelated such that each builds on and improves upon the foundation established by the preceding study. The goals of this dissertation research are as follows:

Study 1. Investigation of baseline neuropsychological subtypes in mild cognitive impairment and their associations with markers of Alzheimer's disease using latent profile analysis.

Study 2. Empirical characterization of longitudinal neuropsychological subtypes in mild cognitive impairment using serial latent profile analysis and multi-year, demographically-corrected norms with embedded practice effects.

Study 3. A latent transition model examining the likelihood of class change in neuropsychological subtypes of mild cognitive impairment.

Study 1: Investigation of baseline neuropsychological subtypes in mild cognitive impairment and their associations with markers of Alzheimer's disease using latent profile analysis

Study 1 will capture neuropsychological heterogeneity among ADNI participants diagnosed with conventional MCI using latent profile analysis to determine baseline empirical subtypes. Inclusion of visuoconstructional measures will allow for the identification of potentially unique cognitive phenotypes.

Specific Aim 1. Investigate unique MCI subtypes at baseline across four neurocognitive domains (visuoconstructional ability, language, attention/executive function, and episodic memory) using LPA.

Aim 1, Hypothesis 1. The optimal LPA solution will generate five classes: four subgroups similar in size and neuropsychological profile to the Edmonds et al. (2015) study and a small, fifth subtype predominantly characterized by visuoconstructional impairment.

Aim 1, Hypothesis 2: Visuoconstructional deficits will be present in a class analogous to the dysexecutive MCI subgroup from Edmonds et al. (2015), thus representing a subtype with "mixed" neuropsychological impairment.

Exploratory Aim: Evaluate LPA class differences on exploratory outcomes of cerebrospinal fluid and genetic AD biomarkers, longitudinal outcome, and other ADNI measures.

Study 2: Empirical characterization of longitudinal neuropsychological subtypes in mild cognitive impairment using serial latent profile analysis and multi-year, demographically-corrected norms with embedded practice effects

Study 2 will examine the stability and consistency of MCI subtypes across 24-months. Separate LPAs will be conducted on baseline, 12-, and 24-months neuropsychological performance to establish a measurement model at each time point. Additionally, study 2 will advance the methods of study 1 by improving upon the normative procedures used to obtain standardized neuropsychological scores. First, the visuoconstructional measures will be excluded in Study 2 due to non-normally distributions, a small range of possible scores, ceiling effects, and poor discrimination normal older adults and individuals with MCI (Eppig et al., 2017). An adaptation of the Heaton, Miller, Taylor & Grant (2004) standardization methodology will be used to produce multi-year, demographically-corrected T-scores based on the normative data of robust normal controls, which should minimize the impact of skewed distributions and outliers. Practice effects will also be accounted for within the standardized scores as the robust normal controls will have also completed all neuropsychological tests at each time point. Reliable change classifications will also be developed based on robust normal control normative data to facilitate comparison of performance over time between the LPA classes.

Specific Aim 1. Develop multi-year neuropsychological norms with demographic correction and embedded practice effects by adapting a previously established, systematic method of standardization (Heaton, Miller, Taylor & Grant, 2004).

Aim 1, Hypothesis 1. Consistent with Heaton, Miller, Taylor & Grant (2004), T-scores in the robust normal control sample will demonstrate several desirable psychometric properties: a mean of 50 and standard deviation of 10, a frequency distribution that does not differ from the Gaussian curve,

minimal outliers, no difference in "impaired" scores (i.e., less than 1 standard deviation below the mean) scores than the expected percentage across tests and time points, and no significant interactions in the final scores between age and education.

Specific Aim 2. Examine exploratory LPA measurement models with the standardized T-scores to longitudinally characterize the stability and consistency of neuropsychological MCI subtypes at baseline, 12-, and 24-months follow-up.

Aim 2, Hypothesis 1: Consistent with study 1, it is hypothesized that 3 classes will consistently emerge as the optimal solution at each time point, with analogous amnestic impairment, multi-domain impairment, and cognitively normal neuropsychological subtypes.

Aim 2, Hypothesis 2: The cognitively normal class will maintain a similar proportional size at all three time points. The amnestic impairment class will shrink in proportional size and the multi-domain impairment class will increase in proportional size over the 24-months.

Specific Aim 3: Investigate change in neuropsychological performance over time between analogous classes from baseline to 12-months, 12-months to 24-months, and baseline to 24-months.

Aim 3, Hypothesis 1: The cognitively normal class will maintain statistically equivalent mean neuropsychological performance over the 24-months, while the amnestic and multi-domain impairment classes will demonstrate mean neuropsychological decline over each time period.

Aim 3, Hypothesis 2: The cognitively normal class will have the smallest proportion of individuals that demonstrate a reliable decrease across all neuropsychological measures over time and the largest percentage with reliable increase. The multi-domain impairment class will have the largest proportion of individuals that demonstrate a reliable decrease over time and the smallest percentage with a reliable increase. The proportion of individuals in the amnestic impairment class with reliable changes will fall between the other two classes.

<u>Study 3: A latent transition model examining the likelihood of class change in neuropsychological subtypes of mild cognitive impairment</u>

Although LTA has been used to model neuroimaging changes in older adults (Gu, 2016), no research has employed LTA to examine transitions between empirical MCI subtypes over time. Therefore,

study 3 will use latent transition analysis (LTA) to investigate changes in class membership of MCI subtypes over 24-months. Study 3 will directly build on study 2, employing the same participants and normative methods, and the LPAs from study 2 will be tested for measurement invariance to facilitate LTA computation. LTA will determine transition probabilities from baseline to 12-months, and 12-months to 24-months. A second-order effect (i.e., direct effect of baseline classification on 24-month classification) and the influence of AD-risk factor covariates (i.e., AD cerebrospinal fluid and genetic biomarkers, functional ability) will be added to the LTA to determine their effects on the likelihood of transition.

Specific Aim 1. Establish measurement invariance between the LPAs from study 2 at baseline, 12-months, and 24-months.

Aim 1, Hypothesis 1. The cognitively normal class will demonstrate fully invariant neuropsychological means for all measures across the three time points.

Aim 1, Hypothesis 2. The amnestic and multi-domain classes will demonstrate partial measurement invariance such that scores in some cognitive domains will decline over the 24-months while others remain stable. Nonetheless, the overall neuropsychological interpretation of their profile will remain equivalent between time points.

Specific Aim 2. Model the probability of changes in class membership from baseline to 12-months and 12-months to 24-months using LTA.

Aim 2, Hypothesis 1: Participants in the cognitively normal and multi-domain classes will be very likely to remain within their respective classes over the 24-months with minimal class change.

Aim 2, Hypothesis 2: Participants in the amnestic impairment class will demonstrate a greater probability of transition to the multi-domain impairment class from baseline to 12-months, and 12- to 24-months than the cognitively normal class.

Specific Aim 3: Determine the influence of AD-risk factor covariates on transition probabilities.

Aim 3, Hypothesis 1: Covariates will increase the probability of transition from the amnestic to the multi-domain impairment class, and the cognitively normal to the amnestic impairment class among individuals with AD-positive risk factors (i.e., functional impairment, apolipoprotein ε4+, AD-cerebrospinal fluid biomarker positivity, greater A/T/N positivity).

Practical Considerations and Clinical Relevance

A significant consideration of this dissertation research involves the practical and clinical relevance of the research. Specifically, the goals, methods, and analyses outlined in this dissertation were selected in part for their potential clinical application to practicing neuropsychologists. For instance, the neuropsychological measures are used as individual indicators in the LPAs and LTA, rather than constructing latent cognitive factors and subsequently performing factor mixture modeling (Clark, Muthén, et al., 2013). Although the latter may be a more statistically sophisticated analysis, neuropsychologists rarely interpret factor scores in clinical practice (cf. WAIS-4 factor indices). However, they do synthesize a variety of test scores (usually representing several cognitive domains) into a meaningful neuropsychological profile. Additionally, a single factor score may obscure interesting patterns among test variables, such as a retrieval or encoding deficit determined by delayed recall and recognition on memory assessment.

Clinical considerations also influenced the selection of LTA to model change over time. Other analyses, such as growth mixture models, evaluate change through the use of latent intercept and slope variables (Ram & Grimm, 2009). This allows researchers to model classes of longitudinal trajectories, which is a potentially interesting and useful analysis. However, clinical neuropsychologists do not typically establish trajectories of patient performance; only in the last decade or so have some begun to employ reliable change indices on repeat assessments. In practice, changes are assessed by clinically comparing an individual's baseline pattern of performance to a repeat evaluation alongside considerations of possible practice effects, and judgments about the possibility of future decline are often based on these profiles. Therefore, LTA was chosen for its ability to model changes in class membership over time, as these likelihoods are potentially valuable information for a clinician attempting to predict future patterns of performance based on an individual's current cognitive profile. In sum, this dissertation uses LPA and LTA over other latent mixture models to facilitate clinical interpretation and increase the applicability to practicing neuropsychologists.

Study 1, in part, is a reprint of the material as it appears in the *Journal of the International Neuropsychological Society, (2017), 23*(7), 564-576. Eppig, Joel S.; Edmonds, Emily C.; Campbell,

Laura; Sanderson-Cimino, Mark; Delano-Wood, Lisa; and Bondi, Mark W. for the Alzheimer's Disease Neuroimaging Initiative, Cambridge University Press. The dissertation author was the primary investigator and author of this paper.

III. Methods

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org. Research was conducted in accordance with the Declaration of Helsinki and the current study approved by the University of California San Diego IRB.

Study 1: Investigation of baseline neuropsychological subtypes in mild cognitive impairment and their associations with markers of Alzheimer's disease using latent profile analysis.

Participants

Participants included 825 individuals diagnosed with MCI and 260 healthy elderly participants. MCI was diagnosed at a screening evaluation using conventional diagnostic criteria, as operationalized by ADNI (Petersen et al., 2010): 1) Subjective memory complaint; 2) Mini-Mental State Examination (MMSE) score greater than or equal to 24; 3) Global Clinical Dementia Rating Scale (CDR) score of 0.5; 4) Impairment on WMS-R Logical Memory-II Story A Recall (WMS-R LM II) after education adjustment; and 5) Intact global cognition and preserved activities of daily living/ instrumental activities of daily living. MCI participants were required to fall within the demographic boundaries of the elderly normative control group, resulting in the exclusion of 19 individuals due to age (i.e., >90 or <60). The final sample consisted of 806 MCI participants. Healthy elderly control subjects (n=260) were required to have complete data on the neuropsychological variables examined and remain cognitively intact upon longitudinal re-evaluation (follow-up range: 1–7 years). Table 2 provides demographic information on these "robust" normal control participants and the entire MCI sample for descriptive purposes.

Table 2: Demographic characteristics of the mild cognitive impairment sample and robust normal controls in study 1.

	Age (years)	Education (years)	Gender	GDS	MMSE at Screening
MCI (*n*=806)	73.90 (6.94)	15.95 (2.81)	40.0% F	1.65 (1.42)	27.58 (1.81)
rNC (*n*=260)	75.25 (5.62)	16.20 (2.68)	48.8% F	0.62 (0.06)	29.05 (1.17)

Note: Data summarized as mean (standard deviation), unless otherwise noted. GDS = Geriatric Depression Scale; MMSE = Mini-Mental State Examination; MCI = Mild Cognitive Impairment; rNC = robust normal controls; F = Female.

Neuropsychological Measures

Eight neuropsychological variables were selected from seven cognitive tests in ADNI's neuropsychological battery. These variables were balanced across the domains of visuoconstructional ability (Mini-Mental State Examination [MMSE] Pentagons & Clock Drawing Test [CDT]); language (Animal Fluency & 30-item Boston Naming Test [BNT]); attention/executive function (Trail Making Test [TMT], Part A & TMT, Part B); and episodic memory (Rey Auditory Verbal Learning Test [AVLT] Delay Free Recall & AVLT Recognition). These specific neuropsychological test variables were selected from available ADNI measures as they were administered across all three ADNI phases and represent well-researched assessments in older adults that are commonly employed and easily interpreted in clinical practice (Lezak, 2012). WMS-R Logical Memory was not selected for the test corpus due its primary use in MCI diagnosis, thereby circumventing criterion contamination.

MMSE Pentagons

Raw MMSE baseline data were obtained via the ADNI website and participant copies of the interlocking pentagons were re-coded using an 8-point error scoring system previously published by Jefferson et al. (2002). This scoring system was chosen to increase the possible range (i.e., 0 to 8 points vs. the standard 0 or 1 scoring system) and minimize potential ceiling effects. Additionally, past research by Jefferson et al. (2002) has shown differential performance in patients with cortical vs. subcortical neurodegenerative disorders using this 8-point scoring system. Errors include 1) size distortion, 2) number of figures, 3) improper pentagon intersection, 4) tremor/segmentation, 5) absence of five angles,

6) significant rotation, 7) interminable motor perseveration, and 8) pull-to-stimulus. For further information and operational definitions of the scoring system, please refer to Jefferson et al. (2002).

Two raters were trained on the 8-point scoring system and established reliability on a randomly selected subset (*n*=54) of MMSE pentagons from the ADNI sample. After establishing satisfactory reliability (*single measure intra-class correlation:* 0.906, 95% CI: 0.838 – 0.945; *range of kappa values for individual error types:* 0.673 – 1.000) each rater was randomly assigned half of the remaining MMSE pentagons for recoding with the 8-point error scoring system.

MMSE pentagon drawings could not be retrospectively obtained via archives for 17.7% of our MCI sample. According to ADNI representatives, data were missing due to technical problems with raw file upload rather than lack of administration or inability to complete the test. Missing values analysis indicated that MMSE pentagons were not missing completely at random (Little's MCAR test: $\chi^2(7)=22.156$, $p=0.002$) when evaluated with the other 7 neuropsychological variables. However, original MMSE pentagon scores (0 or 1) were available in ADNI for all MCI participants. These original scores and the 8-point error scoring system were significantly correlated with a medium effect ($r= –0 .387$, $p<0.001$), supporting their use as a reasonable proxy to examine the missing data. The proportion of individuals with correct versus incorrect original scores did not differ ($\chi^2(1)= 2.519$, $p=0.112$) by presence (n=663; Correct: 87.6%, Incorrect: 12.4%) or absence (n=143; Correct: 92.3%, Incorrect: 7.7%) of raw files. Therefore, raw files were not absent because of poor performance secondary to underlying disease etiology and were assumed missing at random (MAR).

Clock Drawing Test

Clock drawing to command and copy was administered and scored according to ADNI procedures (Alzheimer's Disease Neuroimaging Initiative, 2008; Goodglass & Kaplan, 1983). Briefly, participants were instructed on command to "draw the face of a clock showing the numbers and two hands set to ten after eleven" on blank paper. The participant was then presented a response form with the model clock at the top and requested to "copy this clock (point to the model) in the space provided below".

Clock drawings to command and copy were each scored using the same 0 to 5-point scale. Clock scoring criteria as outlined in the ADNI-2 Procedures Manual (ADNI, 2008) include 1) approximately circular, 2) symmetry of number placement, 3) correctness of numbers, 4) presence of two hands, and 5) presence of two hands set to ten after eleven. Individual command and copy scores were combined to produce an overall Clock Drawing Test total score (0 – 10). This total score was selected for the current analysis rather than separate command and copy scores to maximize the range of possible performance while minimizing any potential ceiling effects. For further information on clock drawing administration and scoring criteria please refer to the ADNI-2 Procedures Manual: http://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf

Transformations and Normative Standardization

The distribution of each neuropsychological variable was examined for non-normality within the sample of robust normal control participants. Each variable was investigated using the ladder function in Stata version 12, which utilizes a chi-square test to determine if and what type of transformation is most appropriate (Tukey, 1977). Animal fluency; TMT, Part A; and TMT, Part B were identified with skew and kurtosis that would significantly benefit from application of the square-root, logarithm-10, and inverse square-root functions, respectively, to improve normality. The remaining five neuropsychological variables (i.e., CDT, MMSE pentagons, BNT, AVLT Recall, and AVLT Recognition) did not significantly benefit from any transformation and therefore retained their identity distributions.

Following application of transformations, standardized regression-based (SRB) formulas were used to generate normative data for each neuropsychological variable based on robust normal control performance. Age, education, and gender were included to account for potential demographic effects, and regression formulas based on these beta coefficients were used to calculate the predicted performance of each MCI participant on all eight neuropsychological variables. This predicted score was then used to obtain a z-score reflecting an MCI subject's degree of impairment on each variable:

$$\textbf{Formula 1: } \text{z-score} = \frac{\text{Observed Score} - \text{Predicted Score}}{\text{Standard Error of the Estimate}}$$

The corresponding z-score formulas and the adjusted R^2 for each regression are displayed in Table 3.

Table 3: Standardized regression formulas based on robust normal control performance in study 1.

Variable	z-score standardized regression formula	Adj. R^2
MMSE Pentagons	[Raw Score – (0.149 + (0.021 x Age) + (-0.021 x Education) + (0.135 x Gender)]/1.004	0.011
Clock Drawing Test	[Raw Score – (11.330 + (-0.043 x Age) + (0.087 x Education) + (-0.012 x Gender)]/0.860	0.125
√ (Animal Fluency)	[Raw Score – (4.690 + (-0.019 x Age) + (0.074 x Education) + (0.093 x Gender)]/0.576	0.119
BNT	[Raw Score – (29.084 + (-0.041 x Age) + (0.153 x Education) + (-0.572 x Gender)]/2.016	0.072
Lg_{10}(TMT, Part A)	[Raw Score – (1.137 + (0.006 x Age) + (-0.006 x Education) + (-0.001 x Gender)]/0.130	0.071
1/√(TMT, Part B)	[Raw Score – (0.187 + (-0.001 x Age) + (0.002 x Education) + 0.004 x Gender)]/0.020	0.138
AVLT Recall	[Raw Score – (12.511 + (-0.106 x Age) + (0.164 x Education) + (1.809 x Gender)]/3.625	0.043
AVLT Recognition	[Raw Score – (13.676 + (-0.019 x Age) + (0.029 x Education) + (0.553 x Gender)]/2.287	0.004

Note: Adj. R^2 = adjusted R^2; SEE = standard error of the estimate; MMSE = mini-mental state examination; BNT= 30-item boston naming test; Lg10 = logarithm 10; TMT = trail making test; AVLT = rey auditory verbal learning test.

Distal Outcome Variables

Distal outcome variables of interest included demographics, ADNI diagnostic measures, biological and genetic markers, longitudinal clinical outcome, and ADNI phase at time of enrollment (ADNI-1, ADNI-GO, ADNI-2). Diagnostic measures used by ADNI to originally identify MCI included WMS-R LM-II score, CDR sum of boxes, MMSE, and the Functional Activities Questionnaire (FAQ). Biological markers were available on 52.4% of MCI (n=422) and 55.0% (n=143) of robust normal control participants; markers included AlzBio3 cerebrospinal fluid (CSF) immunoassay concentrations of total tau (tau), tau phosphorylated at amino acid-181 ($pTau_{181}$), Amyloid-$beta_{1-42}$ ($A\beta_{1-42}$), and the ratio of $pTau_{181}$ to $A\beta_{1-42}$. Subjects were classified according to CSF concentration thresholds (tau: >93 pg/mL; $pTau_{181}$: >23 pg/mL; $A\beta_{1-42}$: <192 pg/mL; $pTau_{181p}/A\beta_{1-42}$ ratio: >0.10) previously established for the Alzbio3 assay to maximize sensitivity and specificity of autopsy confirmed AD (Shaw et al., 2009). Apolipoprotein E (*APOE*) ε4 allele frequency was accessible for 98.8% of MCI participants (n=796) and 100% of the robust normal control participants, and was included in the current study as a genetic susceptibility marker of AD. Longitudinal clinical outcome was available on 93.8% of MCI participants (n=756), with average follow-up of 28.7 months. Variables included type of clinical conversion (progression to dementia, remain stable MCI, or reversion to normal) and the associated number of months to conversion.

Latent Profile Analysis

Data preparation (descriptive statistics, regression, and formatting for import into MPlus) were conducted in SPSS version 22. The ladder command in Stata version 12 was utilized to determine the benefit of transformations on the normality of neuropsychological variables in robust normal controls. All multivariate analyses were performed in MPlus version 7.3.

Latent profile analysis (LPA) was conducted using SRB z-scores of the eight neuropsychological variables as indicators of class membership. Models with two to eight latent classes were evaluated and maximum likelihood estimation with robust standard errors was used in LPA model estimation. Unavailable MMSE pentagons (82% covariance coverage) were assumed missing at random (MAR) in the model. All LPA's were initially performed with the default number of random starts, which were subsequently increased twice (100, 25; and 500, 100) to ensure reproduction of global maxima and protect against misidentification of an erroneous local maxima (Hipp & Bauer, 2006). All LPA results were unchanged after increasing random starts.

Determination of the best-fitting LPA is an iterative process, comparing a model with *k* latent classes to *k*-1 classes until obtaining an optimal solution. Multiple indicators of model fit are useful to determine the best number of latent classes; however, LPA lacks a gold standard and requires consideration of these indices in conjunction with model parsimony and meaningful theoretical interpretation (Berlin, Williams & Parra, 2014; Roesch et al., 2010). Study 1 considered three comparative fit indices: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and sample-size adjusted Bayesian Information Criterion (sBIC), with the smallest values indicating the best-fitting model. In addition, the Vuong-Lo-Mendell-Rubin adjusted Likelihood Ratio Test (VLMR-LRT) and the Bootstrap Likelihood Ratio Test (BLRT) were used to compare the model with *k* latent classes to the *k*-1 class solution; statistical significance ($p<0.05$) suggests the *k* class model is a better fit than *k*-1 classes. Additional statistics used to identify suitable model fit include entropy, an aggregate index of posterior probabilities that reflects the overall precision with which subjects were correctly classified (Berlin, Williams & Parra, 2014; Roesch et al., 2010) and the number of classes containing <5% of the overall sample size, an indicator of potential data over-extraction (Berlin, Williams & Parra, 2014; Roesch, et al.,

2010). Monte Carlo simulation studies using a variety of sample sizes suggest the sBIC, BLRT, and entropy are the most robust fit indices (Berlin, Williams & Parra, 2014; Nylund et al., 2007; Roesch et al., 2010; Tein, Coxe & Cham, 2013). Finally, LPA solutions were evaluated for model parsimony, data over-extraction, and meaningful theoretical interpretation based on previous research.

After selection of the optimal LPA, distal outcome variables were examined between latent classes within the structural equation modeling (SEM) framework. This method is preferential over subject assignment to most likely latent class membership and subsequent ANOVA comparisons; analyzing distal outcomes within SEM models classification uncertainty in statistical comparisons, generating accurate standard errors and reducing biased inferences (Asparouhov & Muthén, 2015; Bray et al., 2015). In the current study the 3-step BCH method (Bakk & Vermunt, 2015; Bolck, Croon, & Hagenaars, 2004; Vermunt, 2010) was employed for continuous distal outcome variables, while the DCAT command was utilized with categorical distal outcome variables (Lanza et al., 2013). The former uses a weighting procedure to account for classification error, while the latter treats distal outcomes as a form of covariate. Asparouhov & Muthén (2015) have demonstrated that these methods are the preferable approaches for continuous and categorical distal outcomes, respectively, due to their satisfactory estimation of standard error, resistance to class shifts, and minimal bias. MPlus performs parameter comparisons on all measures using the Wald chi-square test (Asparouhouv & Muthén, 2007) and statistical significance was set at α=0.005 in study 1 to control for Type-I errors.

Study 2: Empirical characterization of longitudinal neuropsychological subtypes in mild cognitive impairment using serial latent profile analysis and multi-year, demographically-corrected norms with embedded practice effects.

Participants

1,625 individuals enrolled in ADNI through December 2016 were selected for inclusion in studies 2 and 3 due to a baseline diagnosis of MCI (n= 982) or classification as a normal elderly control (n= 643). MCI was diagnosed at a screening evaluation using conventional diagnostic criteria, as operationalized by ADNI (ADNI, 2008; Petersen et al., 2010): 1) Subjective memory complaint beyond what one would expect for age, verified by a study partner; 2) Mini-Mental State Examination (MMSE) score greater than

or equal to 24; 3) Global Clinical Dementia Rating Scale (CDR) score of 0.5, with a score of 0.5 in the Memory Box; 4) Abnormal memory function, operationalized by performing below an education-adjusted cut-score on WMS-R Logical Memory-II Story A Recall (WMS-R LM II); and 5) Preservation of global cognition and functional performance such that Alzheimer's disease (AD) could not be diagnosed. Normal elderly control participants (NC) were defined by ADNI (ADNI, 2008) as 1) Free of memory complaints, beyond what one would expect for age; 2) Normal memory function, operationalized by scoring above an education adjusted cut-off on WMS-R LM II; 3) MMSE greater than or equal to 24; 4) CDR score of 0, with a score of 0 in the Memory Box; 5) Cognitively normal, defined as the absence of significant impairment in cognition or activities of daily living. Additionally, ADNI requires the following of all participants: 1) Hachinski score less than or equal to 4; 2) Geriatric Depression Scale less than 6; 3) Between the ages of 55 and 90 at the time of study enrollment; 4) Completion of 6 years of education or a sufficient work history to rule out intellectual disability; 5) Visual and auditory acuity adequate for neuropsychological testing; 6) Fluency in English or Spanish; 7) Good general health, with no neurological, psychiatric, or other major medical/ systemic illnesses expected to interfere with the study; 8) Women whom are 2 years post-menopausal or surgically sterile; 9) Not using exclusionary medications as specified by ADNI (ADNI, 2008); and 10) Not currently enrolled in another trial or study.

In studies 2 and 3, the robust normal control (rNC) sample was utilized to develop standardized T-scores for normative reference at baseline, 12-months, and 24-months, as well as z-scores indicating reliable change between the three time points. Robust normal control samples, defined as individuals that remain normal over time, have been shown to improve the accuracy of normative data in older adults compared to the inclusion of normal individuals with incipient MCI and/or dementia (Harrington et al., 2017; Holtzer et al., 2008; De Santi et al., 2008). rNC participants in studies 2 and 3 were required to retain an ADNI classification of normal (follow-up range: 1-10 years) throughout their entire ADNI enrollment, as updated through December 2016. Of the initial 643 normal control participants identified by ADNI, 385 met this "robust" normal criteria. Additionally, inclusion criteria for studies 2 and 3 required rNCs to have at least 2 years of follow-up (n= 291) and complete data on all neuropsychological measures examined at baseline, 12-months, and 24-months. Thus, the final sample of rNCs consisted of

284 individuals with complete scores on all neuropsychological variables of interest, across all three time points. Figure 1 provides a visual representation of the inclusion criteria, associated sample size, and selection of robust normal controls for studies 2 and 3.

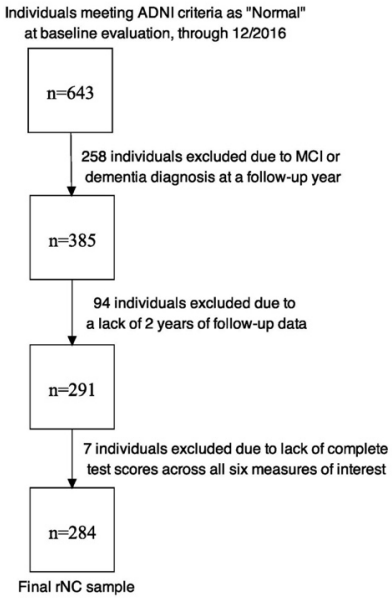

Figure 1: Derivation of robust normal control sample for studies 2 and 3. ADNI = Alzheimer's Disease Neuroimaging Initiative; MCI = mild cognitive impairment; rNC = robust normal control.

MCI participants in studies 2 and 3 were required to fall within the demographic boundaries of the robust normative control group. Thus, of the initial 982 MCI participants, 144 were excluded due to age at baseline (i.e., >89 or <60 years) and 1 due to low education (i.e., <6 years). MCI participants were also required to have at least one neuropsychological score available in each cognitive domain investigated (language, attention/ executive functions, episodic memory), resulting in the exclusion of an additional 12 individuals. These inclusion criteria led to a final sample of 825 MCI participants at baseline. Individuals with a baseline MCI diagnosis, available follow-up, and at least one neuropsychological score in each cognitive domain of interest at the respective time point were included for analysis at 12-months (n= 751) and 24-months (n= 639). Figure 2 provides a visual representation of the inclusion criteria, associated sample size, and selection of MCI participants in studies 2 and 3.

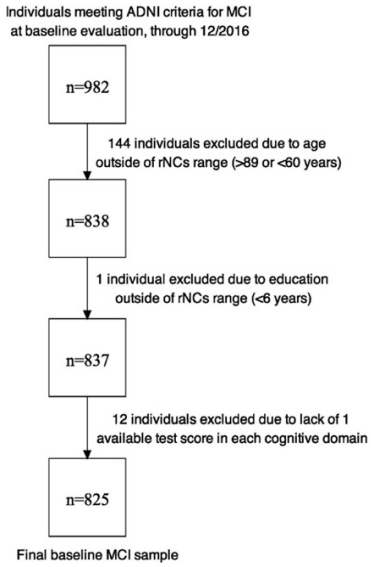

Figure 2: Derivation of the mild cognitive impairment sample for studies 2 and 3. ADNI = Alzheimer's Disease Neuroimaging Initiative; MCI = mild cognitive impairment; rNC = robust normal control.

Although MCI participants were allowed to have one missing score in each of the three cognitive domains of interest to maximize sample size across time points, the vast majority had all neuropsychological measures available at baseline (n= 813, 98.5%), 12-months (n= 743, 98.9%), and 24-months (n= 621, 97.2%). Table 4 provides demographic information on the final sample of rNC and MCI participants at baseline, 12-months, and 24-months in studies 2 and 3 for descriptive purposes.

Table 4: Demographic characteristics of the mild cognitive impairment and robust normal control samples at baseline, 12-months, and 24-months in studies 2 and 3.

Time Point	Group	Age (years)	Education (years)	Gender	GDS	MMSE at Screening
Baseline	rNCs (*n*=284)	74.49 (5.60)	16.43 (2.73)	47.5% F	0.72 (1.08)	29.09 (1.16)
	MCI (*n*=825)	73.84 (6.92)	15.91 (2.83)	40.4% F	1.66 (1.42)	27.59 (1.80)
12-Months	rNCs (*n*=284)	75.49 (5.60)	16.43 (2.73)	47.5% F	0.91 (1.40)	28.96 (1.36)
	MCI (n=751)	74.82 (6.93)	15.98 (2.82)	40.2% F	1.88 (1.93)	27.07 (2.60)
24-Months	rNCs (*n*=284)	76.49 (5.60)	16.43 (2.73)	47.5% F	1.01 (1.35)	29.05 (1.17)
	MCI (n=639)	75.56 (6.88)	16.03 (2.75)	40.2% F	1.90 (2.02)	26.40 (3.44)

Note: Data summarized as mean (standard deviation), unless otherwise noted. GDS = geriatric depression scale; MMSE = mini-mental state examination; rNCs= robust normal controls; MCI = mild cognitive impairment; F = female.

Neuropsychological Measures

Six neuropsychological variables were selected from five cognitive tests in ADNI's neuropsychological battery for studies 2 and 3. These variables were balanced across three cognitive domains: language (Animal Fluency & 30-item BNT), attention/executive function (TMT, Part A & TMT, Part B), and episodic memory (AVLT Delay Free Recall & AVLT Recognition). These specific neuropsychological test variables were selected from available ADNI measures as they were administered across all three ADNI phases, represent well-researched assessments in older adults, are commonly employed in clinical practice (Lezak, 2012), and have been used extensively in past MCI classification research (Bondi et al., 2014; Edmonds et al., 2015; Edmonds et al., 2016; Eppig et al., 2017). Additionally, these variables provide a reasonable range of possible scores with desirable psychometric characteristics in a normative sample; visuoconstructional measures from study 1 failed to meet this criterion, and were therefore excluded from studies 2 and 3. WMS-R Logical Memory was also excluded due to its primary use in ADNI's MCI diagnosis, thereby circumventing criterion contamination. Heaton et al. (2004) qualitative classifications (above average, average, below average, mild impairment, mild-to-moderate impairment, moderate impairment, moderate-to-severe impairment, severe impairment) were used as descriptors of neuropsychological performance based on the corresponding T-score range (55+, 50-54, 45-49, 40-44, 35-39, 30-34, 25-29, 20-24, 0-19).

Multi-Year Normative Standardization

Standardization procedures for baseline, 12-months, and 24-months neuropsychological performance were adapted from the methodology established by Heaton et al. (1991), and later replicated and advanced in other normative studies (Cysique et al., 2011; Gladsjo et al., 1999; Heaton et al., 2004; Norman et al., 2011). The first step in normative standardization involved conversion of neuropsychological raw scores into scaled scores with a mean of 10 and standard deviation of 3, based on quantiles from the rNC frequency distribution. This conversion helps non-normal and skewed raw scores conform to a Gaussian distribution, attenuates the effect of outliers, places all test scores on the same metric, and ensures that lower scores always indicate worse performance across tests (Cysique et

al., 2011; Gladsjo et al., 1999; Heaton et al., 1991; Heaton et al., 2004; Norman et al., 2011). This procedure was performed separately for baseline, 12-months, and 24-months using the respective rNC test distribution at each time point. As all rNCs had complete neuropsychological data across all years and tests were administered in a standardized fashion with identical test-retest intervals, this conversion also automatically embedded practice effects within the scaled score distributions. This novel design supports instantaneous comparison of performance across time and circumvents the need for clinical judgement or post-hoc statistical methods to correct for practice effects (Cysique et al., 2011; Heaton et al., 2001).

Following conversion of raw to scaled scores, multiple fractional polynomial regression based on rNC performance was employed to generate demographically-corrected normative data using the methodology of Heaton et al. (2004) for each neuropsychological variable at each time point. The Royston and Altman fractional polynomial method was used to compare all combinations of demographic predictors with a set of exponents (–2, –1, –0.5, 0, 0.5, 1, 2, 3) to determine the optimal model fit (Cysique et al., 2011; Heaton et al., 2004; Norman et al., 2011). Demographic predictors included age in years, sex (dummy coded: 0 = male, 1 = female), and education as a semi-continuous variable in the multiple fractional polynomial regressions. Due to the very small sample of rNC participants at each level of education below 12 years (6 years: n= 1, 7 years: n= 1, 9 years: n= 1, 10 years: n=3), individuals with 11 or fewer years of education were re-coded into a single category (≤11 years). However, the original, continuous ADNI classification was maintained for participants with 12 through 20 years of education. Stratified age by education cell counts are presented in Table 5.

Table 5: Robust normative sample cell counts by age & education in study 2.

		Education (years)					
		<=11	12	13–15	16	17–18	19+
Age (years)	60–64	0	1	0	2	4	1
	65–69	1	3	8	11	6	14
	70–74	3	9	25	30	25	17
	75–79	1	6	19	17	13	20
	80–84	0	1	3	10	10	14
	85–89	1	1	4	0	1	3

Upon generation of the multiple fractional polynomial regressions, the predicted scaled score of each participant on all six neuropsychological variables at baseline, 12-months, and 24-months was

calculated from these models for a total of 18 predicted scaled scores per individual. The predicted scaled score was then inserted in the following formula to obtain a T-score reflecting a subject's degree of impairment on each neuropsychological variable, at each time point:

$$\textbf{Formula 2: } \mathrm{T-score} = \left[\left(\frac{Observed\ Scaled\ Score - Predicted\ Scaled\ Score}{Standard\ Error\ of\ the\ Estimate\ of\ the\ Regression}\right)\times 10\right] + 50$$

These normative procedures were chosen to generate T-scores with several desirable psychometric properties in elderly normal individuals: 1) a mean of 50 and standard deviation of 10; 2) generally normalized distributions with minimal outliers; 3) no associations with age, education, or sex; 4) and equivalent interpretation of performance across both tests and time (Heaton et al., 2004). These psychometric properties were evaluated for all 18 T-scores in the rNC sample to determine the efficacy of the normative methods. In addition to reporting means and standard deviations, the chi-square test was used to evaluate differences in the proportion of rNCs with an "impaired" score (i.e., less than 1 standard deviation below the mean) versus the expected 15.9% based on a normal distribution, and the Kolmogorov-Smirnov test was used to determine if each rNC T-score distribution differed from normality. Linear regressions with age, education, and sex predicting each T-score were used to check for any residual association between T-scores and demographics, and the age x education interaction term was subsequently included to ensure that associations did not differ across levels of age and education. Statistical significance was set at α=0.01 to control for Type-I errors given the replication of each above analysis across the 18 T-scores (Heaton et al., 2004). Multiple fractional polynomial regression was calculated in Stata version 12; subsequent T-score analyses were conducted in SPSS version 24.

Reliable Change Classification

In addition to the multi-year normative standardization, reliable change z-scores were calculated based on rNC performance to provide a metric of individual change across time. Past research has demonstrated that reliable change indices are useful for determining if a person's change in performance is beyond the level expected in a normal population (Attix et al., 2009; Cysique et al., 2011; Duff, 2012; Heaton et al., 2001). Several different reliable change formulas have been created to account for various test-retest issues (Duff, 2012; Stein, Luppa, Brähler, König, & Riedel-Heller, 2010). Study 2 utilized a

regression-based method that accounts for test-retest reliability and regression to the mean, and has proven effective at detecting reliable change in prior neuropsychological research (Attix et al., 2009; Cysique et al., 2011; Duff, 2012).

Three separate linear regressions were performed for each neuropsychological variable: baseline T-score predicting 12-months T-score, baseline T-score predicting 24-month T-score, and 12-month T-score predicting 24-month T-score. Demographics were already accounted for in the T-scores and therefore not included in reliable change regressions. The predicted T-scores for 12- and 24-months among the three regressions was used to obtain a z-score with the following formula:

$$\textbf{Formula 3: } \text{z–score} = \frac{Observed\ T–Score - Predicted\ T–Score}{Standard\ Error\ of\ the\ Estimate\ of\ the\ Regression}$$

The resultant z-score indicated the amount of change a participant demonstrated at 12- and 24-months, relative to the typical change in rNCs with similar performance at baseline or 12-months. These z-scores were used to categorize an individual with an increase, decrease, or stable performance based on the clinical standard of a 90% confidence interval (Cysique et al., 2011; Duff, 2012; Heaton et al., 2001). Thus, a z-score above 1.645 (corresponding to the top 5% of the rNC distribution) would indicate a "significant increase", a z-score below -1.645 (corresponding to the bottom 5% of the rNC distribution) a "significant decrease", and "stable" being equivalent to a z-score equal to or in-between 1.645/ -1.645. Chi-square tests were used to determine if the proportion of rNC participants in each category significantly differed from the expected outcomes of 5%, 90%, and 5% across each measure. Statistical significance was again set at α=0.01 to control for Type-I errors. Reliable change analyses were conducted in SPSS version 24.

Serial Latent Profile Analyses

Three exploratory latent profile analyses (LPA) were conducted separately at baseline, 12-months, and 24-months to investigate measurement models of the neuropsychological performance among participants diagnosed with MCI at baseline. The six neuropsychological T-scores were used as indicators of class membership for the serial LPAs, with baseline scores indicating the baseline LPA, 12-month scores indicating the 12-month LPA, and 24-month scores indicating the 24-month LPA. Models

with two to eight latent classes were evaluated and full information maximum likelihood estimation with robust standard errors was used in LPA model estimation. All LPA's were initially performed with 100 random sets of starting values for the initial stage and 20 optimizations for the final stage (STARTS= 100, 20; Muthén & Muthén, 2012). Analyses were re-run with increased starts (200, 40; 500, 100; etc.) until the best log-likelihood was reproduced at least twice to ensure reproduction of global maxima and protect against misidentification of an erroneous local maxima (Hipp & Bauer, 2006; Muthén & Muthén, 2012). Figure 3 provides a visual representation of the three proposed latent models.

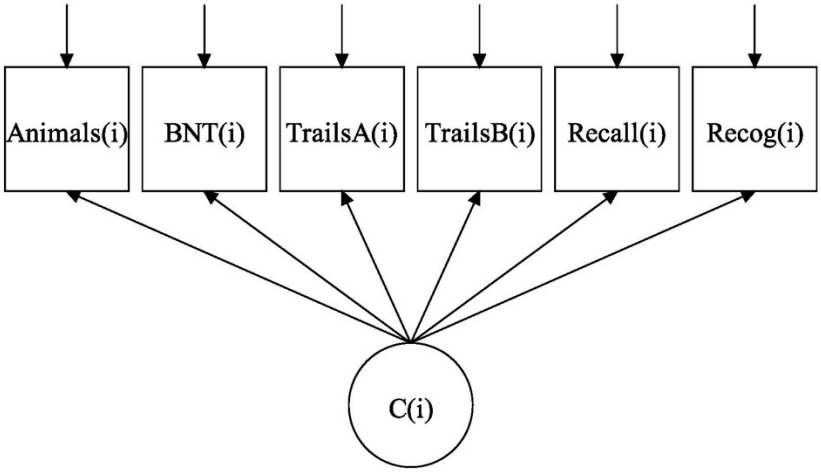

Figure 3: A graphical model of the latent profile analyses. Observed indicator variables of the categorical latent class variable at the i^{th} time point (i= baseline, 12-months, or 24-months). Animals = animal fluency T-score; BNT = 30-item boston naming test T-score; TrailsA = trail making test, part A T-score; TrailsB = trail making test, part B T-score; Recall = rey auditory verbal learning test recall T-score; Recog = rey auditory verbal learning test recognition T-score; C = categorical latent class variable.

Determination of the best-fitting LPA is an iterative process, comparing a model with *k* latent classes to *k*-1 classes until obtaining an optimal solution. Multiple indicators of model fit are useful to determine the best number of latent classes; however, LPA lacks a gold standard and requires consideration of these indices in conjunction with model parsimony and meaningful theoretical interpretation (Berlin, Williams & Parra, 2014; Roesch, Villodas & Villodas, 2010). Study 2 considered three comparative fit indices: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and sample-size adjusted Bayesian Information Criterion (sBIC), with the smallest values indicating the best-fitting model. In addition, the Vuong-Lo-Mendell-Rubin adjusted Likelihood Ratio Test (VLMR-LRT) and the Bootstrap Likelihood Ratio Test (BLRT) were used to compare the model with *k* latent classes to the

k-1 class solution; statistical significance ($p<0.05$) suggests the *k* class model is a better fit than *k*-1 classes. Additional statistics used to identify suitable model fit include entropy, an aggregate index of posterior probabilities that reflects the overall precision with which subjects were correctly classified (Berlin et al., 2014; Roesch et al., 2010) and the number of classes containing <5% of the overall sample size, an indicator of potential data over-extraction (Berlin et al., 2014; Roesch, et al., 2010). Monte Carlo simulation studies using a variety of sample sizes suggest the sBIC, BLRT, and entropy are the most robust fit indices (Berlin et al., 2014; Nylund et al., 2007; Roesch et al., 2010; Tein, Coxe & Cham, 2013). Additionally, past neuropsychological LPA research has shown that the VLMR-LRT may also be particularly useful in identification of optimal model fit when using cognitive test indicators (Eppig et al., 2017; Dajani, Llabre, Nebel, Mostofsky & Uddin, 2016; Köhler et al., 2013; Morin & Axelrod, 2017). Finally, LPA solutions were evaluated for model parsimony, data over-extraction, and meaningful theoretical interpretation based on previous research models. Data preparation (descriptive statistics, linear regressions, and formatting for import into MPlus), was conducted in SPSS version 24. All multivariate analyses were performed in MPlus version 7.4.

Parameter and Outcome Comparisons

After selection of the optimal LPA at each time point, additional analyses involving parameter comparisons (i.e., class means) and distal outcomes were examined within the mixture modeling framework. This method is preferential over subject assignment to most likely latent class membership and subsequent ANOVA comparisons; analyzing parameters and distal outcomes within the mixture model includes classification uncertainty in statistical comparisons, generating accurate standard errors and reducing biased inferences (Asparouhov & Muthén, 2015; Bray et al., 2015). With respect to parameter comparisons, mean neuropsychological performance was compared within each analogous cognitive class across time points and between cognitive classes within the same time point using the Wald chi-square test (Asparouhouv & Muthén, 2007). The Wald chi-square test was also used to evaluate differences in distal outcomes via mixture regression analysis. Age and education were compared between classes at each respective time point using the 3-step BCH method (Bakk & Vermunt, 2015; Bolck, Croon, & Hagenaars, 2004; Vermunt, 2010) and sex was examined between classes using the

DCAT command (Lanza et al., 2013). Two additional categorical distal outcomes were evaluated at 12- and 24-months using the DCAT command: concurrent ADNI diagnosis (i.e., normal, MCI, or dementia) and reliable change classification (i.e., decline, no change, improvement). Asparouhov & Muthén (2015) have demonstrated that these respective methods are the preferable approaches for continuous and categorical distal outcomes due to their satisfactory estimation of standard error, resistance to class shifts, and minimal bias. Statistical significance for omnibus-level comparisons was set at α=0.05, and Bonferroni correction was used to control for Type-I errors among multiple post-hoc comparisons.

Study 3: A latent transition model examining the likelihood of class change in neuropsychological subtypes of mild cognitive impairment.

Participants

Participants included all 825 individuals from study 2 that were diagnosed with baseline MCI using ADNI criteria (ADNI, 2008; Petersen et al., 2010). Follow-up data was available on 91.0% of the baseline sample at 12-months (n=751) and 77.5% of baseline subjects at 24-months (n=639). MCI participants were enrolled between the ages of 60 and 89 at their baseline visit, had 6 or more years of education, and had at least one neuropsychological score available in each cognitive domain at all time points. The vast majority of MCI participants possessed a complete set of scores across all neuropsychological measures of interest at baseline (n=813, 98.5%), 12-months (n=743, 98.9%), and 24-months (n=621, 97.2%).

Neuropsychological Measures

Neuropsychological measures were identical to the six ADNI variables used in study 2 (Animal Fluency; 30-item Boston Naming Test [BNT]; Trail Making Test [TMT], Part A; TMT, Part B; Rey Auditory Verbal Learning Test [AVLT] Delay Free Recall; and AVLT Recognition). Raw neuropsychological scores were converted to demographically-adjusted T-scores with embedded practice effects using an adaption of the Heaton et al. (2004) methodology. A detailed review of these standardization procedures as well as characteristics of the robust normal control sample used to establish the normative data is described under the methods of study 2.

Covariate Variables

Covariate variables of interest included baseline functional ability, Apolipoprotein E (APOE) genotype, and baseline AD cerebrospinal fluid biological markers. All variables were dummy coded (unless otherwise specified) based on previously established relationships in the literature with AD dementia (Saunders et al., 1993; Schindler et al., 2018; Teng et al., 2010). Baseline functional ability was measured with the Functional Activities Questionnaire (FAQ) and classified as intact (total FAQ<6) or impaired (total FAQ≥6) using a cut-score previously shown to maximize the sensitivity and specificity for discriminating between MCI and very mild AD dementia (Teng et al., 2010). Baseline FAQ score was available in 99.5 – 99.6% of individuals at baseline (n=822), 12-months (n=748), and 24-months (n=636). In the overall baseline MCI sample, 78.7% an FAQ<6 and 21.2% had a score ≥6. APOE ε4 allele frequency was included as a genetic marker of AD (Saunders et al., 1993) and classified as ε4-positive (ε4+) or negative (ε4-). APOE genotype was available in 99.5% of individuals at baseline (n=821), and 100% at 12- (n=751) and 24-months (n=639). In the overall baseline MCI sample, 49.1% were ε4-negative and 50.9% were ε4-positive.

Baseline AD biological markers were available in a large subsample of the 825 MCI participants, with the three markers of interest available for 71.6% of individuals at baseline (n=591), 73.6% at 12-months (n=553), and 74.0% at 24-months (n=473). Baseline cerebrospinal (CSF) concentrations of total tau (tTau), tau phosphorylated at amino acid-181 ($pTau_{181}$), and amyloid-$beta_{1-42}$ ($A\beta_{1-42}$) were measured with the automated Elecsys immunoassays, which have been recently shown to decrease variability and improve consistency of CSF results compared to prior methods (Bittner et al., 2016; Schindler et al., 2018). Subjects were identified as AD positive (AD+) or negative (AD-) for each biomarker according to CSF concentration thresholds (AD+ tTau >242 pg/mL; AD+ $pTau_{181}$ >19.2 pg/mL; AD+ $A\beta_{1-42}$ <1098 pg/mL) to maximize the sensitivity and specificity of Elecsys immunoassays with amyloid-positive PET scans (Schindler et al., 2018). These variables were subsequently examined as covariates in the LTA analysis. In the overall baseline MCI sample, 69.0% were classified as $pTau_{181}$ positive, 57.0% as tTau positive, and 68.0% as $A\beta_{1-42}$ positive.

In addition to AD-positivity/negativity on the three separate CSF biomarkers, each individual was classified according to the Amyloid/Tau/Neurodegeneration (A/T/N) scheme proposed by Jack et al. (2016), using $A\beta_{1\text{-}42}$ for Amyloid, $pTau_{181}$ for Tau, and tTau for Neurodegeneration. Thus, each participant was classified into one of eight possible A/T/N combinations, which was represented as one categorical A/T/N variable. A/T/N classification of the total baseline MCI sample with available AD-CSF biomarkers (n=591) is presented in Table X. Of the eight possibilities, three categories (A–/T+/N–, A–/T–/N+, A+/T–/N+) contained cell counts so small (<5% of the total sample) that these individuals were excluded from subsequent analyses. The final sample consisted of 566 participants with baseline A/T/N classification that was used as an LTA covariate.

Table 6: A/T/N classification across the total baseline MCI sample with AD-CSF biomarkers (n=591).

Classification	A–/T–/N–	A+/T–/N–	A–/T+/N–	A–/T–/N+	A+/T+/N–	A+/T–/N+	A–/T+/N+	A+/T+/N+
Sample Size	n=91	n=89	n=22	n=3	n=52	n=0	n=73	n=261
Percentage	15.4%	15.1%	3.7%	0.5%	8.8%	0.0%	12.4%	44.2%

Note: A/T/N = amyloid/tau/neurodegeneration scheme; MCI = mild cognitive impairment; AD = Alzheimer's disease; CSF = cerebrospinal fluid.

Latent Transition Analysis

Data preparation (descriptive statistics, dummy coding, and formatting for import into MPlus) was conducted in SPSS version 24. All multivariate analyses were performed in MPlus version 7.4. Latent transition analysis is a multi-step process that involves 1) Examining descriptive statistics and data patterns of selected indicators, 2) Investigating separate LPAs at each time point and exploring alternative measurement models, 3) Determining measurement invariance across the longitudinal LPAs (i.e., same number of classes, equivalent class means or conditional item probabilities within an analogous class), 4) Specifying the latent transition model without higher-order effects or covariates, and 5) Including higher-order effects and/or covariates in the LTA model (Nylund, 2007). LTA steps 1 and 2 were conducted in study 2, and the parameter means from these results were used as starting values in the current LTA model.

In step 3, measurement invariance between the three LPAs was examined by constraining model parameters (i.e. means) and comparing the statistical fit of the constrained versus unrestricted models

(Finch, 2015). Three possibilities may occur in the model: full measurement invariance (i.e., the number of classes and analogous class means are all invariant over time), partial measurement invariance (i.e., the number of classes but only some means within a class are equal over time, while others are freely estimated), and full non-invariance (i.e., no constraints are placed on parameters over time). Although measurement invariance is not required to perform LTA modeling, there are several benefits to a fully (or even partially) invariant model. In addition to establishing equivalent interpretation of longitudinal classes, measurement invariance significantly reduces the computational burden and decreases the likelihood of model non-identification as fewer parameters need to be estimated in the LTA (Collins & Lanza, 2013; Nylund, 2007). Nested LPAs were compared using a likelihood ratio test (LRT) with scaling correction on the model log-likelihoods (Bryant & Satorra, 2012; Muthén & Muthén, 2006; Satorra & Bentler, 2010), such that a significant chi-square ($p<0.05$) indicated worse model fit of the constrained (i.e., non-invariant) LPA. Full measurement invariance was evaluated first and tested within analogous cognitive classes by constraining the LPA means for each respective neuropsychological indicator to be equal across the three time points (e.g., mean Animal Fluency at 12-months in class 1 equal to mean Animal Fluency at baseline in class 1). If full invariance was rejected, partial measurement invariance was then evaluated by constraining only neuropsychological means within a class that were not found to statistically differ ($p\geq0.05$) across all three time points in study 2.

After constraining the necessary means across classes to reflect the appropriate level of measurement invariance, autoregressive relationships between categorical latent variables were specified to determine their relationships over time (step 4). First, the categorical latent variable at 12-months was regressed on the categorical latent variable at baseline; then the categorical latent variable at 24-months was regressed on the categorical latent variable at 12-months. Transition probabilities between these time points were expressed as probabilities in a latent transition matrix to better understand the likely change in class membership for an individual given their initial class membership. LPA class means from study 2 were specified as starting values, with 500 sets of starting values for the initial stage and 100 optimizations for the final stage (STARTS= 500, 100; Muthén & Muthén, 2012). Analyses were re-run with increased starts until the best log-likelihood was reproduced at least twice to

ensure reproduction of global maxima and protect against misidentification of an erroneous local maxima (Hipp & Bauer, 2006; Muthén & Muthén, 2012). All latent transition analyses were performed with full information maximum likelihood estimation with robust standard errors and Bayesian analysis to perform multiple imputation of missing data under the missing at random assumption (Muthén & Muthén, 2012). Figure 4 provides an example visual representation of possible transition probabilities that reflect change in class membership over time.

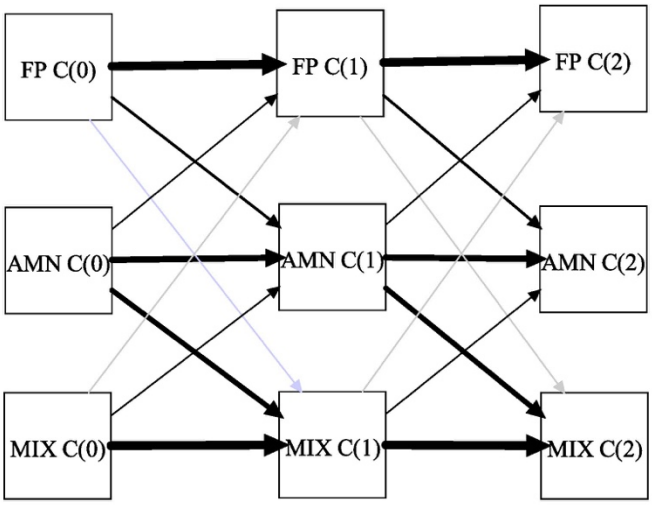

Figure 4: Graphical depiction of transition probabilities between hypothesized latent classes over three time points. Increasing arrow thickness/darkness indicates a greater probability of transition between the two classes at associated time points. FP = false-positive class; AMN = amnesic class; MIX = mixed class; C(0) = class at baseline; C(1) = class at 12-months; C(2) = class at 24-months.

Next, a second-order effect (i.e., categorical latent variable at 24-months regressed on the categorical latent variable at baseline) was added to the prior LTA model that only contained first-order effects (step 5). Class means from the LTA in step 4 were specified as starting values, and the appropriate means were constrained across time to reflect the previously determined level of measurement invariance. Latent transition probabilities resulting from the second-order effect were compared to the transition probabilities from the prior LTA model (i.e., only first-order effects) to examine the lasting direct effect that class membership at baseline has on class membership at 24-months. As an example, a model of an LTA with a second-order effect is presented in Figure 5 below.

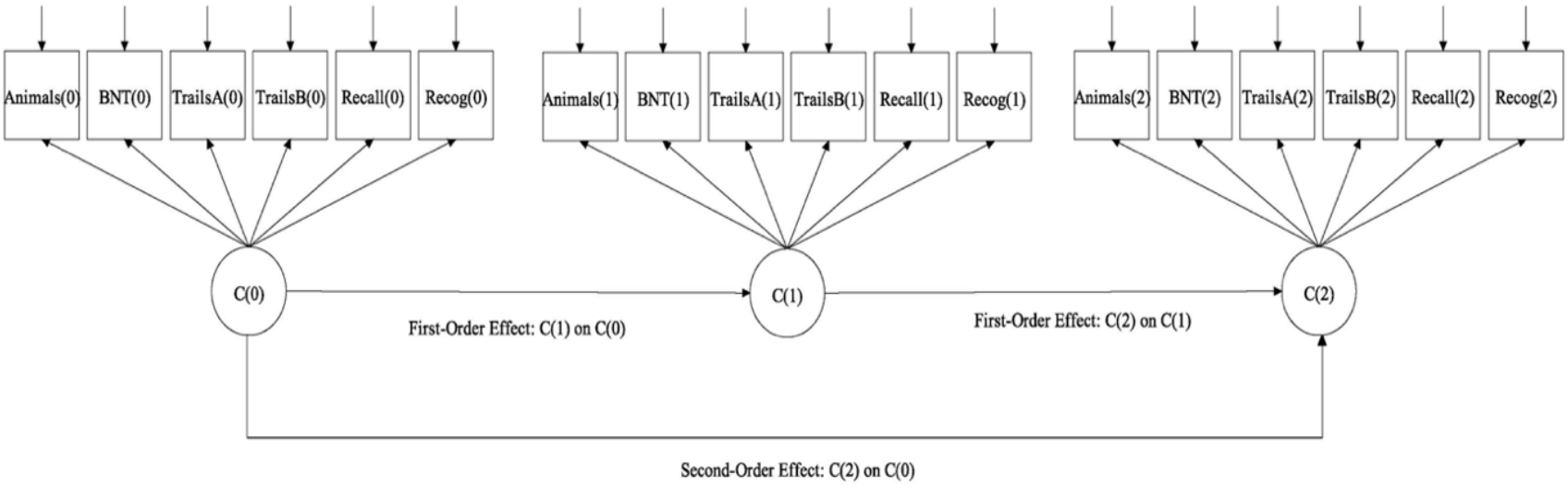

Figure 5: Latent transition analysis with a second-order effect. Animals = Animal Fluency T-score; BNT = 30-item Boston Naming Test T-score; TrailsA = Trail Making Test, Part A T-score; TrailsB = Trail Making Test, Part B T-score; Recall = Rey Auditory Verbal Learning Test Recall T-score; Recog = Rey Auditory Verbal Learning Test Recognition T-score; C = Categorical Latent Class Variable; (0) = variable at baseline; (1) = variable at 12-months; (2) = variable at 24-months.

Finally, covariates were added to the LTA model to investigate effects on latent transition probabilities across time points (step 5). Baseline FAQ, APOE genotype, each baseline AD-CSF biomarker (i.e., tTau, $pTau_{181}$, $A\beta_{1\text{-}42}$), and the A/T/N categorical variable (Jack et al., 2016) were separately modeled on the LTA first-order effects (i.e., 12-months regressed on Baseline, 24-months regressed on 12-months) using the known-class command and Tech 15 output (Muthén & Muthén, 2012). Latent transition probabilities were examined between AD+ and AD- groups for baseline FAQ, APOE genotype, and baseline AD-CSF biomarkers to investigate differences in likelihood of class change in the presence of AD-risk factors. Additionally, the five levels of A/T/N classification (i.e., A–/T–/N–, A+/T–/N–, A+/T+/N–, A–/T+/N+, A+/T+/N+) were evaluated for differential probabilities of class transition. A representation of an example LTA model with a covariate is presented in Figure 6 below.

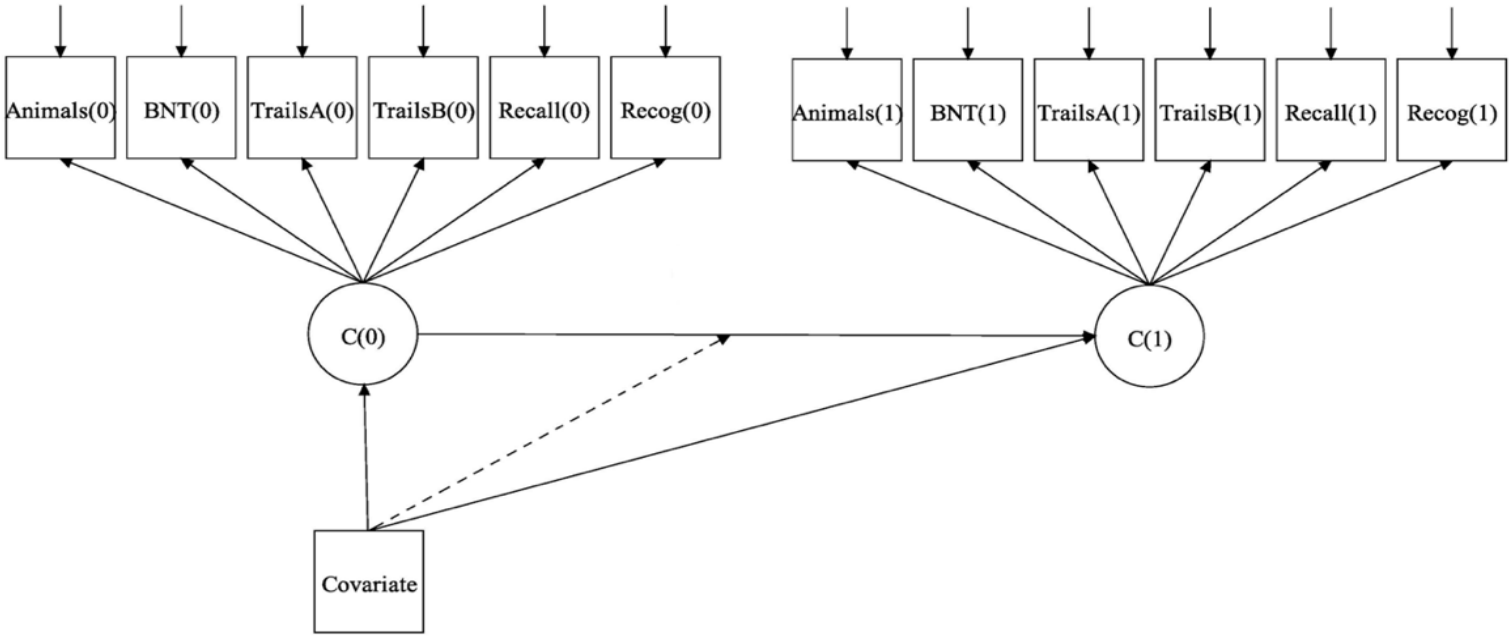

Figure 6: An example of latent transition analysis with a covariate. Animals = Animal Fluency T-score; BNT = 30-item Boston Naming Test T-score; TrailsA = Trail Making Test, Part A T-score; TrailsB = Trail Making Test, Part B T-score; Recall = Rey Auditory Verbal Learning Test Recall T-score; Recog = Rey Auditory Verbal Learning Test Recognition T-score; C = Categorical Latent Class Variable; (0) = Variable at baseline; (1) = Variable at 12-months.

Study 1, in part, is a reprint of the material as it appears in the *Journal of the International Neuropsychological Society, (2017), 23*(7), 564-576. Eppig, Joel S.; Edmonds, Emily C.; Campbell, Laura; Sanderson-Cimino, Mark; Delano-Wood, Lisa; and Bondi, Mark W. for the Alzheimer's Disease Neuroimaging Initiative, Cambridge University Press. The dissertation author was the primary investigator and author of this paper.

IV. Results

Study 1: Investigation of baseline neuropsychological subtypes in mild cognitive impairment and their associations with markers of AIzheimer's disease using latent profile analysis.

Latent Profile Analysis

Two to eight latent class models were tested in study 1. Fit indices and descriptive characteristics for each model are provided in Table 7. AIC, BIC, and sBIC comparative fit indices successively decreased with increasing latent classes; the BLRT showed a similar pattern, with *k* classes always a statistically significant fit compared to *k*-1 classes. These indices failed to clearly converge on an optimal solution, as this trend would presumably continue past eight latent classes and likely result in data over-fitting based on other indicators (Nylund et al., 2007). The VLMR-LRT suggested the 3-class solution as a significantly better fit than 2-classes. However, the 4-class solution (vs. 3-classes) did not result in statistically significant improvement in model fit via the VLMR-LRT. The VLMR-LRT remained non-significant for all subsequent class comparisons (e.g., 5- vs. 4-classes, etc.). Entropy was highest for the 4-class model, though satisfactory (Asparouhov & Muthén, 2014; Tein et al., 2013) and relatively equivalent for the 3-class solutions. The smallest class size for the 3-class solution was 13.2% of all MCI participants; LPA models with 5 or greater classes contained at least one class that was <5% of the overall sample. The 3-class LPA was selected as the optimal solution on the basis of fit indices (e.g., VLMR-LRT), satisfactory entropy, model parsimony, signs of possible data over-fitting with increasing latent classes, and meaningful neuropsychological interpretation of classes.

Table 7: Latent profile analysis comparative fit indices & likelihood ratio tests in study 1.

Number of Classes	AIC	BIC	sBIC	VLMR-LRT	BLRT	Final Log-Likelihood	Entropy	Number of Classes <5%	Smallest Class Size
2	20291.57	20408.87	20329.48	p=0.0265	p<0.0001	-10120.782	0.671	0	39.95%
3*	19982.63	20142.16	20034.19	p=0.0152	p<0.0001	-9957.315	0.773	0	13.15%
4	19820.68	20022.44	19885.90	p=0.2045	p<0.0001	-9867.341	0.823	0	5.83%
5	19710.91	19954.90	19789.77	p=0.0622	p<0.0001	-9803.455	0.764	1	3.60%
6	19560.23	19846.45	19652.74	p=0.1000	p<0.0001	-9719.116	0.771	1	3.85%
7	19502.07	19830.52	19608.23	p=0.4687	p<0.0001	-9681.037	0.789	2	2.85%
8	19435.09	19805.77	19554.89	p=0.3352	p<0.0001	-9638.548	0.781	2	2.48%

Note: *Chosen as best class solution. AIC = akaike information criterion; BIC = bayesian information criterion; sBIC = sample-size adjusted bayesian information criterion; VLMR-LRT = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test; BLRT = bootstrapped likelihood ratio test.

The final 3-class LPA grouped MCI participants into a “mixed” MCI class (n=106, 13.2%), an “amnestic” MCI class (n=455, 56.5%), and an “LPA-derived normal” class (n=245, 30.4%) based upon neuropsychological performance. Final class counts based on most likely class membership is presented in Table 8. Posterior probabilities for correct classification ranged from 0.40 to 1.00 for the mixed MCI class, 0.47 to 1.00 for the amnestic MCI class, and 0.50 to 1.00 for the LPA-derived normal class.

Table 8: Final class counts and proportions for most likely class membership of 3-class LPA in study 1.

	n	Proportion of Total MCI Sample
Mixed MCI Class	106	13.15%
Amnestic MCI Class	455	56.45%
LPA-Derived Normal Class	245	30.40%

Note: LPA = latent profile analysis; MCI = mild cognitive impairment.

Neuropsychological Measures

The mixed MCI class yielded a profile of neuropsychological impairment across all four cognitive domains, ranging from mild-to-moderate to severe deficits. However, performance on the MMSE pentagon test was only low average for the mixed MCI class. The amnestic MCI class demonstrated mild-to-moderate impairment on both measures of episodic memory and average to low average performance across all other cognitive domains. The LPA-derived normal class demonstrated average performance across all neuropsychological tests, despite their original MCI diagnosis. Neuropsychological performance of each class is presented in Figure 2.

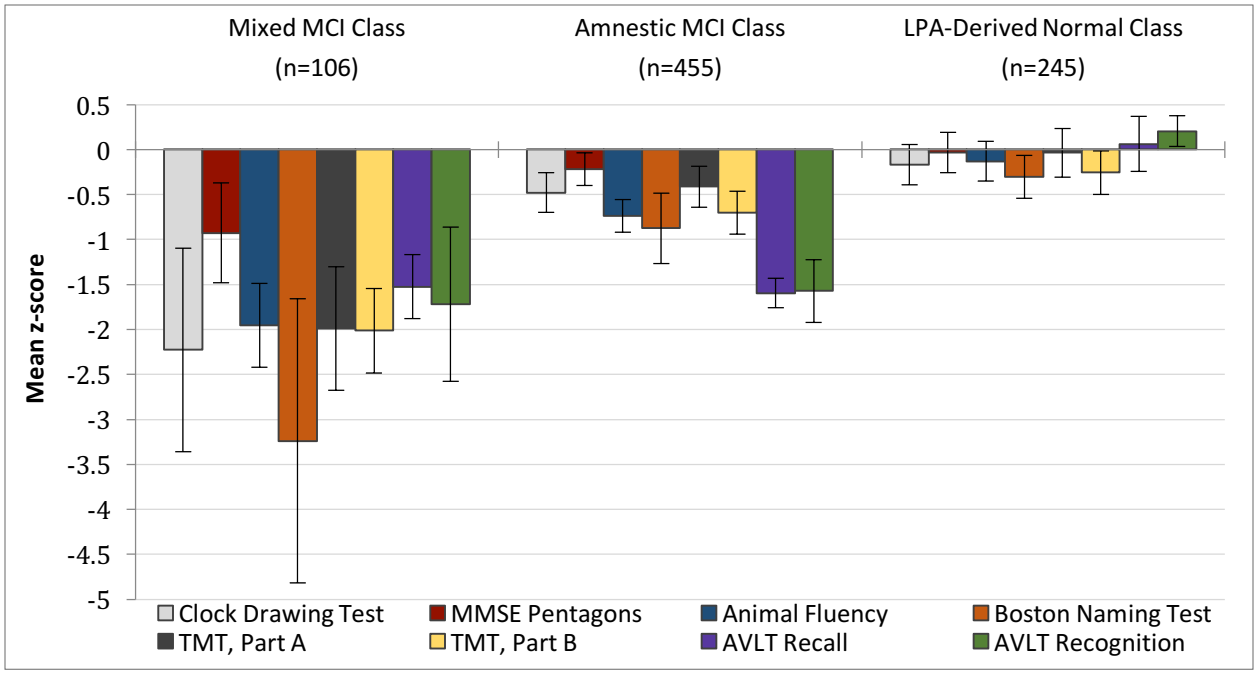

Figure 7: Neuropsychological performance of the latent profile classes in study 1. Error bars denote 99.5% confidence intervals. MCI = mild cognitive impairment; LPA = latent profile analysis; MMSE = mini-mental state examination; TMT = trail making test; AVLT = rey auditory verbal learning test.

Omnibus Wald tests suggested significant differences between classes on every neuropsychological variable (all *p's*<0.001). Post-hoc comparisons indicated the mixed MCI class performed significantly worse (all *p's*<0.001) than both the amnestic MCI and LPA-derived normal classes on all measures of visuoconstructional ability, language, and attention/executive functioning. However, on AVLT Recall and Recognition the mixed MCI class was only significantly worse compared to the LPA-derived normal class ($p< 0.001$). The amnestic MCI class produced significantly lower scores than the LPA-derived normal class on all tests of episodic memory and language, as well as TMT, Part B ($p< 0.001$). There was no statistical difference in performance between the two groups on measures of visuoconstructional ability or TMT, Part A. Differences in neuropsychological performance between classes are presented in Table 9.

Table 9: Neuropsychological performance of the three latent profile classes in study 1.

Variable	Mixed MCI Class	Amnestic MCI Class	LPA-Derived Normal Class	Omnibus Wald χ^2 Test (df)	*p*-value
Visuoconstructional Ability					
MMSE Pentagons	-2.226[2,3] (0.403)	-0.478[1] (0.079)	-0.169[1] (0.080)	$\chi^2(2)=17.829$	$p<0.001$
Clock Drawing Test	-0.927[2,3] (0.198)	-0.217[1] (0.064)	-0.035[1] (0.080)	$\chi^2(2)=27.685$	$p<0.001$
Language					
Animal Fluency	-1.953[2,3] (0.166)	-0.737[1,3] (0.065)	-0.130[1,2] (0.078)	$\chi^2(2)=100.948$	$p<0.001$
BNT	-3.238[2,3] (0.562)	-0.874[1,3] (0.139)	-0.302[1,2] (0.086)	$\chi^2(2)=49.202$	$p<0.001$
Attention/Executive Function					
TMT, Part A	-1.987[2,3] (0.245)	-0.413[1] (0.082)	-0.036[1] (0.097)	$\chi^2(2)=51.823$	$p<0.001$
TMT, Part B	-2.013[2,3] (0.168)	-0.702[1,3] (0.084)	-0.256[1,2] (0.087)	$\chi^2(2)=82.083$	$p<0.001$
Episodic Memory					
AVLT Recall	-1.524[3] (0.126)	-1.596[3] (0.059)	0.061[1,2] (0.034)	$\chi^2(2)=395.429$	$p<0.001$
AVLT Recognition	-1.719[3] (0.306)	-1.572[3] (0.125)	0.205[1,2] (0.060)	$\chi^2(2)=529.387$	$p<0.001$

Note: Data summarized as mean (standard error) in standardized regression-based z-scores. Numbered superscripts denote significant Wald χ^2 test post-hoc differences at $p<0.005$ between each class and the class number indicated (1= mixed MCI class, 2= amnestic MCI class, 3= LPA-derived normal class). MCI = mild cognitive impairment; LPA = latent profile analysis; χ^2 = chi-square; df= degrees of freedom; MMSE = mini-mental state examination; BNT = 30-item boston naming test; TMT = trail making test; AVLT = rey auditory verbal learning test.

Distal Outcomes Variables

Certain variables (e.g., CSF biomarkers, APOE allele, longitudinal outcomes) only included a subset of the total MCI sample; the distribution of these subsamples across the three latent classes is in Table 10.

Table 10: Distribution of subsample variables across the latent classes in study 1.

	MMSE Pentagons		*CSF Biomarkers*		*APOE Alleles*		*Longitudinal Data*	
	n	Proportion of Class	*n*	Proportion of Class	*n*	Proportion of Class	*n*	Proportion of Class
Mixed MCI Class	95	89.62%	52	49.06%	104	98.11%	103	97.27%
Amnestic MCI Class	374	82.20%	237	52.09%	450	98.90%	429	94.29%
LPA-Derived Normal Class	194	79.18%	133	54.29%	242	98.78%	224	91.43%

Note: MMSE = mini-mental state examination; CSF = cerebrospinal fluid; APOE = Apolipoprotein E; MCI= mild cognitive impairment; LPA = latent profile analysis.

Latent class differences on all distal outcome variables are presented in Table 11. Omnibus Wald tests indicated no significant differences between classes on the demographic variables of age, education, gender, or Geriatric Depression Scale (GDS). Significant omnibus differences (all *p's*<0.001) were noted between classes on all ADNI diagnostic measures (i.e., WMS-R LM-II, CDR Sum of Boxes, MMSE, and FAQ). The LPA-Derived normal class performed significantly better on all ADNI diagnostic measures than both the mixed and amnestic MCI class. The amnestic MCI class produced a significantly higher MMSE score and lower CDR Sum of Boxes tally than the mixed MCI class, though no differences between the two classes were noted on WMS-R LM-II or the FAQ.

Table 11: Demographic, diagnostic, genetic, cerebrospinal fluid biomarker, longitudinal, and ADNI phase differences between latent profile classes in study 1.

Category	Variable	Mixed MCI Class	Amnestic MCI Class	LPA-Derived Normal Class	Omnibus Wald χ^2 Test (df)	*p*-value
Demographics	Age	73.72 (0.72)	74.01 (0.35)	73.79 (0.55)	$\chi^2(2)=0.165$	$p=0.921$
	Education	15.77 (0.37)	15.82 (0.15)	16.25 (0.19)	$\chi^2(2)=3.158$	$p=0.206$
	Gender (%)	42.1% F (5.6)	36.6% F (2.5)	45.4% F (4.0)	$\chi^2(2)=3.015$	$p=0.222$
	GDS	1.85 (0.16)	1.54 (0.07)	1.75 (0.11)	$\chi^2(2)=3.690$	$p=0.158$
Diagnostic Measures	WMS-R LM II	4.12[3] (0.359)	4.73[3] (0.18)	8.03[1,2] (0.19)	$\chi^2(2)=178.914$	$p<0.001$
	CDR Sum of Boxes	2.0[2,3] (0.11)	1.58[1,3] (0.05)	1.17[1,2] (0.05)	$\chi^2(2)=63.624$	$p<0.001$
	Baseline MMSE	26.34[2,3] (0.19)	27.41[1,3] (.10)	28.43[1,2] (0.11)	$\chi^2(2)=105.151$	$p<0.001$
	FAQ	4.87[3] (0.54)	3.69[3] (0.23)	1.38[1,2] (0.20)	$\chi^2(2)=75.677$	$p<0.001$
Genetic & CSF Biomarkers	% APOE ε4-positive	61.3%[3] (6.3)	57.8%[3] (3.4)	34.5%[1,2] (3.4)	$\chi^2(2)=30.014$	$p<0.001$
	% high total tau	53.1%[3] (9.2)	42.5%[3] (3.9)	22.6%[1,2] (3.6)	$\chi^2(2)=17.159$	$p<0.001$
	% high $pTau_{181}$	84.2%[3] (6.4)	64.0%[3] (4.4)	33.4%[1,2] (4.3)	$\chi^2(2)=51.233$	$p<0.001$
	% low $A\beta_{1-42}$	81.6%[3] (6.9)	73.0%[3] (3.6)	29.3%[1,2] (4.3)	$\chi^2(2)=72.773$	$p<0.001$
	% high $pTau_{181}/A\beta_{1-42}$ ratio	80.3%[3] (6.9)	76.2%[3] (3.8)	34.6%[1,2] (4.3)	$\chi^2(2)=64.444$	$p<0.001$
Longitudinal Outcome	% progression to dementia	60.0%[2,3] (6.7)	38.3%[1,3] (2.9)	5.8%[1,2] (1.8)	$\chi^2(2)=133.050$	$p<0.001$
	Months until progression	15.85[2,3] (1.58)	23.42[1] (1.46)	51.15[1] (11.25)	$\chi^2(2)=20.338$	$p<0.001$
	% reversion to normal	1.2%[3] (1.2)	0.6%[3] (0.6)	11.2%[1,2] (2.1)	$\chi^2(2)=21.037$	$p<0.001$
	Months until reversion	10.32 (1.71)	18.11 (4.29)	21.45 (3.69)	$\chi^2(2)=8.619$	$p=0.013$
	% stable MCI	38.8%[2,3] (6.7)	61.1%[1,3] (2.8)	83.0%[1,2] (2.6)	$\chi^2(2)=65.159$	$p<0.001$
	Amount of total follow-up (in months)	27.74 (2.37)	29.23 (1.23)	28.02 (1.79)	$\chi^2(2)=0.403$	$p=0.818$
Phase	% from ADNI1	64.3%[3] (5.2)	53.7%[3] (2.8)	27.4%[1,2] (3.2)	$\chi^2(2)=51.420$	$p<0.001$
	% from ADNIGO	0.0%[2,3] (0.0)	9.9%[1,3] (1.8)	28.5%[1,2] (3.1)	$\chi^2(2)=125.303$	$p<0.001$
	% from ADNI2	35.7% (5.2)	36.4% (2.5)	44.1% (3.4)	$\chi^2(2)=3.915$	$p=0.141$

Note: Data summarized as mean or percent of class and (standard error) unless otherwise noted. Numbered superscripts denote significant Wald χ^2 test post-hoc differences at $p<0.005$ between each class and the class number indicated (1=mixed MCI class, 2=amnestic MCI class, 3=LPA-derived normal class). MCI = mild cognitive impairment; LPA = latent profile analysis; χ^2 = chi-square; df= degrees of freedom; GDS = geriatric depression scale; WMS-R LM II = Weschler memory scale-revised logical memory II subtest; CDR = clinical demetia rating scale; MMSE = mini-mental state examination; FAQ = functional activities questionnaire; CSF = cerebrospinal fluid; APOE = Apolipoprotein E; ADNI1 = Alzheimer's Disease Neuroimaging Initiative phase 1; ADNIGO = Alzheimer's Disease Neuroimaging Initiative phase GO; ADNI2 = Alzheimer's Disease Neuroimaging Initiative phase 2.

Significant omnibus differences were also present for all genetic and AD-CSF biomarkers (all *p's*<0.001) available on a subset of the overall sample. A significantly lower proportion of the LPA-derived normal class had the APOE ε4 allele than both other classes (all *p's*<0.001); the mixed MCI and amnestic MCI classes did not differ. A similar pattern emerged for CSF biomarkers: both MCI classes contained a significantly higher percentage of subjects with AD-positive CSF biomarkers (i.e., high total tau, high $pTau_{181}$, low $A\beta_{1\text{-}42}$, and high $pTau_{181}/A\beta_{1\text{-}42}$ ratio) than the LPA-derived normal class (all *p's*<0.003), while the amnestic and mixed MCI classes did not differ. Identical results were obtained upon examination of mean CSF biomarker concentrations, with the exception that post-hoc total tau levels were only a nonsignificant trend between the LPA-derived normal and the mixed MCI class (*p*=0.007). This trend is due to the use of α=0.005 significance level to adjust for multiple comparisons and increased variability of total tau in the mixed class, which produced a larger standard error than the other classes. Additionally, the LPA-derived normal class did not differ from robust normal controls on any of the CSF biomarker concentrations (all p's>0.101). Mean CSF biomarker concentrations between all latent classes as well as robust normal controls are presented in Figure 8.

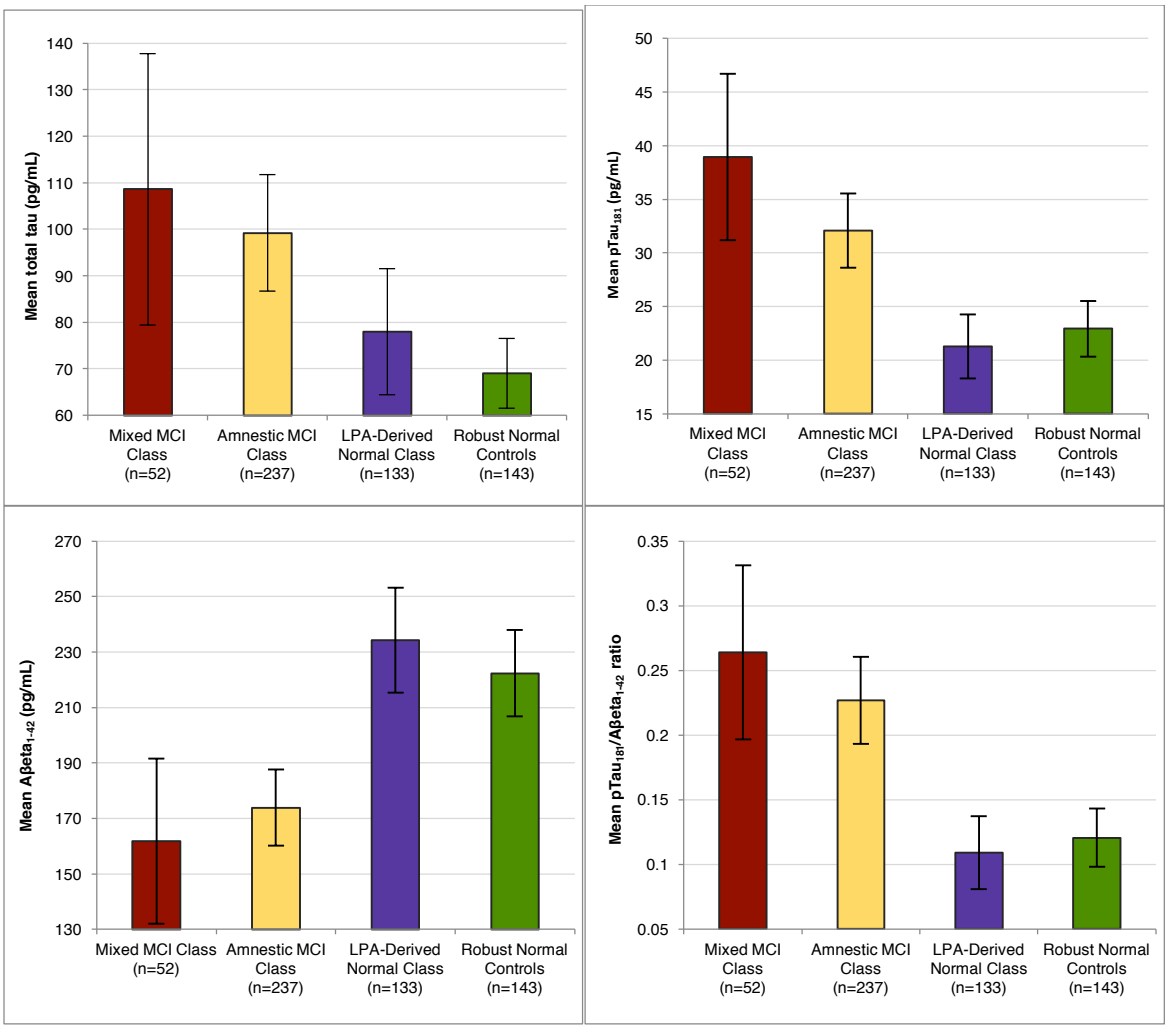

Figure 8: Mean cerebrospinal fluid Alzheimer's disease biomarker concentrations of latent profile classes and robust normal control participants in study 1.
Upper Left Panel: Total tau concentration. *Upper Right Panel:* $pTau_{181}$ concentration.
Lower Left Panel: $A\beta_{1\text{-}42}$ concentration. *Lower Right Panel:* Ratio of $pTau_{181}$ to $A\beta_{1\text{-}42}$. Error bars denote 99.5% confidence intervals. MCI = mild cognitive impairment; LPA = latent profile analysis; $A\beta_{1\text{-}42}$ = amyloid-$beta_{1\text{-}42}$; $pTau_{181}$ = tau phosphorylated at amino acid-181.

With respect to longitudinal outcomes, there was no significant difference between latent classes in amount of available follow-up. However, omnibus differences were noted among the proportion of individuals who progressed to AD diagnoses, reverted to normal, and remained as stable MCI. In particular, a significantly smaller percentage of the LPA-derived normal class progressed to AD than the other classes (all *p's*<0.001). A larger proportion of the LPA-derived normal class also reverted to normal or remained as stable MCI than both other classes (all *p's*<0.001). Compared to the mixed MCI class, the amnestic MCI class had a significantly smaller proportion of individuals who progressed to AD but larger

percentage who remained stable (all *p's*<0.003); no difference was noted in reversion to normal. The mixed MCI class progressed to AD more quickly than both other classes (all *p's*<0.002); no difference was observed in progression time between the amnestic MCI and LPA-derived normal classes. Clinical progression rates for LPA classes are presented in Figure 9.

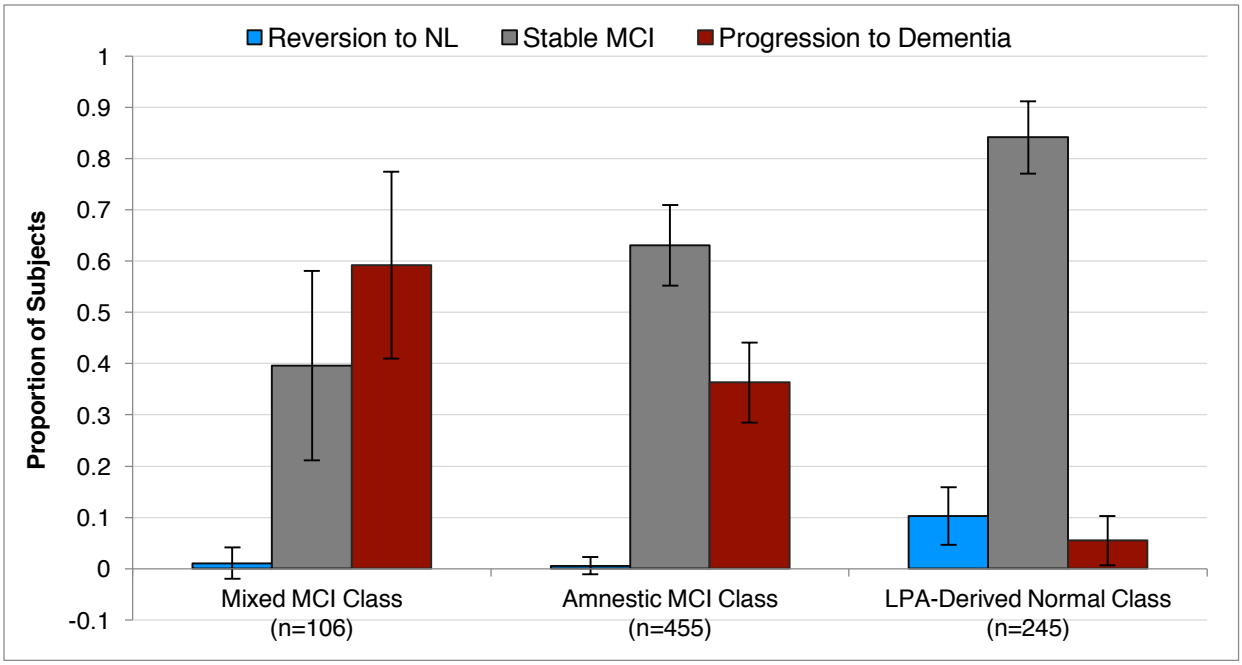

Figure 9: Progression and reversion rates of latent profile classes in study 1. Error bars denote 99.5% confidence intervals. NL = normal; MCI = mild cognitive impairment; LPA = latent profile analysis.

Upon investigation of ADNI enrollment phase, there was no difference in the proportion of individuals recruited during Phase 2. However, significant omnibus differences were noted between the classes for both Phase 1 and ADNI GO. A significantly smaller percentage of the LPA-derived normal class was enrolled during Phase 1 than both other classes (all p's< 0.001); no difference was observed between the mixed and amnestic MCI classes. The opposite trend emerged for ADNI GO, such that a significantly larger percentage of the LPA-derived normal class was enrolled during this phase than both other classes (all *p's*< 0.001). The amnestic MCI class also had a significantly larger proportion of participants recruited during ADNI GO than the mixed MCI class (p< 0.001), as the latter enrolled no individuals in this phase.

Study 2: Empirical characterization of longitudinal neuropsychological subtypes in mild cognitive impairment using serial latent profile analysis and multi-year, demographically-corrected norms with embedded practice effects.

Multi-Year Demographically-Corrected T-scores

The first step to generate the demographically-corrected T-scores involved the conversion of raw scores to scaled scores based on the rNC quantiles at each time point. Conversion tables for the six neuropsychological variables of interest at baseline, 12-months, and 24-months are presented in Tables 12, 13, and 14, respectively.

Table 12: Raw score to scaled score conversions for measures at baseline.

Scaled Score	Animal Fluency	BNT	TMT, Part A	TMT, Part B	AVLT Recall	AVLT Recognition	Percentile
18	38+		0 - 15	0 - 35			>99th
17	34 - 37		16 - 17	36 - 38			99th
16	31 - 33		18	39 - 40	15		98th
15	29 - 30		19 - 20	41 - 44	14		95th
14	27 - 28	30	21 - 23	45 - 51	12 - 13	30	91st
13	25 - 26		24 - 25	52 - 55	11		84th
12	24		26 - 28	56 - 62	10	29	75th
11	22 - 23	29	29 - 30	63 - 71	9	28	63rd
10	20 - 21		31 - 34	72 - 80	7 - 8		50th
9	18 - 19	28	35 - 38	81 - 92	6	27	37th
8	16 - 17	27	39 - 43	93 - 103	5	26	25th
7	15	26	44 - 48	104 - 119	3 - 4	24 - 25	16th
6	13 - 14	25	49 - 51	120 - 164	2	22 - 23	9th
5	12	24	52 - 60	165 - 197	1	20 - 21	5th
4	11	22 - 23	61 - 68	198 - 229	0	19	2nd
3	9 - 10	21	69 - 73	230 - 299		16 - 18	1st
2	0 - 8	0 - 20	74 - 150	300		0 - 15	<1st

Note: BNT = 30-item boston naming test; TMT = trail making test; AVLT = rey auditory verbal learning test.

Table 13: Raw score to scaled score conversions for measures at 12-months.

Scaled Score	Animal Fluency	BNT	TMT, Part A	TMT, Part B	AVLT Recall	AVLT Recognition	Percentile
18	38+		0 -16	0 - 31			>99th
17	35 - 37		17	32 - 34			99th
16	32 - 34		18	35 - 40	15		98th
15	30 - 31		19 - 20	41 - 44	14		95th
14	28 - 29		21	45 - 48	13	30	91st
13	26 - 27	30	22 - 25	49 - 54	12		84th
12	24 - 25		26 - 27	55 - 60	11	29	75th
11	22 - 23		28 - 29	61 - 67	9 - 10	28	63rd
10	20 - 21	29	30 - 32	68 - 74	8	27	50th
9	19		33 - 36	75 - 89	6 - 7	26	37th
8	17 - 18	28	37 - 41	90 - 100	5	25	25th
7	16	27	42 - 45	101 - 127	3 - 4		16th
6	14 - 15	25 - 26	46 - 48	128 - 162	1 - 2	23 - 24	9th
5	13	24	49 - 61	163 - 199	0	21 - 22	5th
4	12		62 - 68	200 - 236		18 - 20	2nd
3	11	23	69 - 75	237 - 299		16 - 17	1st
2	0 - 10	0 - 22	76 - 150	300		0 - 15	<1st

Note: BNT = 30-item boston naming test; TMT = trail making test; AVLT = rey auditory verbal learning test.

Table 14: Raw score to scaled score conversions for measures at 24-months.

Scaled Score	Animal Fluency	BNT	TMT, Part A	TMT, Part B	AVLT Recall	AVLT Recognition	Percentile
18	38+		0 - 15	0 - 33			>99th
17	36 - 37		16	34 - 36			99th
16	32 - 35		17	37 - 39	15		98th
15	30 - 31		18 - 19	40 - 44	14		95th
14	27 - 29		20 - 22	45 - 48	13	30	91st
13	26	30	23 - 24	49 - 54	12		84th
12	24 - 25		25 - 26	55 - 62	11	29	75th
11	22 - 23	29	27 - 28	63 - 68	10	28	63rd
10	20 - 21		29 - 31	69 - 76	8 - 9		50th
9	18 - 19		32 - 35	77 - 90	6 - 7	27	37th
8	16 - 17	28	36 - 39	91 - 109	5	26	25th
7	15	27	40 - 43	110 - 128	3 - 4	25	16th
6	14	25 - 26	44 - 49	129 - 152	1- 2	24	9th
5	12 - 13	24	50 - 62	153 - 192	0	21 - 23	5th
4	11	23	63 - 70	193 - 249		19 - 20	2nd
3	10	20 - 22	71 - 90	250 - 299		17 - 18	1st
2	0 - 9	0 - 19	91 - 150	300		0 - 16	<1st

Note: BNT = 30-item boston naming test; TMT = trail making test; AVLT = rey auditory verbal learning test.

Age, education, and sex were then used as predictors of these scaled scores in a series of multiple fractional polynomial regressions. The Royston and Altman algorithm determined that linear relationships (e.g., power of 1) between the predictors and scaled scores provided the best fit for each model. All models were significant at $p \leq 0.01$, and the overall variance accounted for ranged from 2.8% – 19.0% across models. Combined demographics accounted for more than 10% of the variance in Animal Fluency, Trail Making Test Part A, and Trail Making Test Part B across all time points, whereas they explained less than 10% of the variance in the 30-item Boston Naming Test, AVLT Recall, and AVLT Recognition. The predicted scaled scores and standard error of the model estimates were then applied in Formula 2 to calculate demographically-corrected T-scores for the six measures across the three time points. Formulas derived from the multiple fractional polynomial regressions, adjusted R^2 for each overall model, and the partial r^2 accounted for by each demographic predictor is presented in Table 15.

Table 15: T-score formulas with embedded practice effects derived from robust normal control performance using multiple fractional polynomials.

Time Point	Measure	Multiple Fractional Polynomial T-score Formula	Adjusted R^2	Age/Education/Sex Partial r^2
Baseline	Animal Fluency	[Scaled Score.0 – (14.153 + (-0.119 x Age.0) + (0.293 x Education) + (0.334 x Sex))/ 2.802] x 10 + 50	.100	.054/.064/.003
	BNT	[Scaled Score.0 – (12.749 + (-0.057 x Age.0) + (0.116 x Education) + (-0.871 x Sex))/ 2.874] x 10 + 50	.061	.012/.018/.021
	TMT, Part A	[Scaled Score.0 – (19.949 + (-0.154 x Age.0) + (0.103 x Education) + (0.321 x Sex))/ 2.833] x 10 + 50	.084	.085/.008/.003
	TMT, Part B	[Scaled Score.0 – (22.125 + (-0.196 x Age.0) + (0.159 x Education) + (0.532 x Sex))/ 2.688] x 10 + 50	.153	.144/.021/.009
	AVLT Recall	[Scaled Score.0 – (15.684 + (-0.120 x Age.0) + (0.190 x Education) + (1.203 x Sex))/ 2.856] x 10 + 50	.092	.053/.027/.040
	AVLT Recognition	[Scaled Score.0 – (11.9353 + (-0.0488 x Age.0) + (0.0854 x Education) + (1.3633 x Sex))/ 2.7771] x 10 + 50	.054	.010/.006/.053
12-Months	Animal Fluency	[Scaled Score.12 – (16.772 + (-0.142 x Age.12) + (0.258 x Education) + (-0.248 x Sex))/ 2.786] x 10 + 50	.115	.076/.051/.002
	BNT	[Scaled Score.12 – (15.602 + (-0.108 x Age.12) + (0.200 x Education) + (-0.570 x Sex))/ 2.571] x 10 + 50	.094	.052/.037/.011
	TMT, Part A	[Scaled Score.12 – (20.259 + (-0.172 x Age.12) + (0.164 x Education) + (0.756 x Sex))/ 2.752] x 10 + 50	.127	.110/.022/.017
	TMT, Part B	[Scaled Score.12 – (23.325 + (-0.206 x Age.12) + (0.152 x Education) + (0.207 x Sex))/ 2.731] x 10 + 50	.156	.152/.019/.001
	AVLT Recall	[Scaled Score.12 – (1.844 + (-0.075 x Age.12) + (0.221 x Education) + (1.334 x Sex))/ 2.760] x 10 + 50	.081	.023/.038/.052
	AVLT Recognition	[Scaled Score.12 – (14.081 + (-0.081 x Age.12) + (0.131 x Education) + (0.794 x Sex))/ 2.655] x 10 + 50	.046	.029/.015/.020
24-Months	Animal Fluency	[Scaled Score.24 – (13.704 + (-0.112 x Age.24) + (0.301 x Education) + (0.328 x Sex))/ 2.765] x 10 + 50	.101	.049/.069/.003
	BNT	[Scaled Score.24 – (14.8480 + (-0.085 x Age.24) + (0.151 x Education) + (-0.517 x Sex))/ 2.566] x 10 + 50	.058	.033/.021/.009
	TMT, Part A	[Scaled Score.24 – (21.371 + (-0.188 x Age.24) + (0.194 x Education) + (0.604 x Sex))/ 2.817] x 10 + 50	.140	.124/.029/.011
	TMT, Part B	[Scaled Score.24 – (20.423 + (-0.195 x Age.24) + (0.267 x Education) + (0.895 x Sex))/ 2.636] x 10 + 50	.190	.147/.061/.026
	AVLT Recall	[Scaled Score.24 – (12.341 + (-0.075 x Age.24) + (0.205 x Education) + 1.046 x Sex))/ 2.853] x 10 + 50	.058	.022/.031/.031
	AVLT Recognition	[Scaled Score.24 – (11.259 + (-0.045 x Age.24) + (0.143 x Education) + (0.883 x Sex))/ 2.805] x 10 + 50	.028	.008/.016/.023

Note: All model *p's*≤0.001 except AVLT Recognition at 24-Months: *p*=0.01. Robust normal control sample: n=284. BNT= 30-item Boston Naming Test; TMT = Trail Making Test; AVLT = Rey Auditory Verbal Learning Test; Scaled Score.0 = Scaled score at baseline; Scaled Score.12 = Scaled Score at 12-months; Scaled Score.24 = Scaled Score at 24-months.

The demographically-corrected T-scores were subsequently examined within the rNC sample (n=284) to evaluate their psychometric properties. Across the 18 T-scores, means ranged from 49.88 to 50.11, and standard deviations from 9.90 to 10.01. The percentage of scores in the "impaired" range for each of the 18 T-scores – defined as less than 1 standard deviation below the mean (i.e., T<40) – fell from 12.3% to 17.6% in the rNC sample, and there were no significant differences from the percentage of individuals expected to fall within that range (i.e., 15.9%) on the normal curve (all *p*'s ≥0.104). Kolmogorov-Smirnov tests used to assess the normality of each T-score distribution in the rNC sample indicated that distributions for Animal Fluency, TMT Part A, TMT Part B, and AVLT Recall at each of the

three time points did not differ from a normal curve (all *p*'s>0.013). However, the distributions for BNT and AVLT Recognition at baseline, 12-months, and 24-months were significantly different compared to a Gaussian distribution (all *p*'s<0.001). As expected, separate linear regressions using age, education, and sex as demographic predictors for each of the 18 T-scores in the rNC sample was not significant (all *p*'s≥0.996). Partial variance explained by the separate demographic predictors for each T-score was also non-significant (all *p*'s>0.807), and the corresponding partial r^2 associations were all less than 0.03%. Inclusion of the age x education interaction term in each of these 18 models was not significant at α=0.01 (12-month Animal Fluency: *p*=0.029, all other model *p*'s> 0.072), indicating that the lack of association between the demographic predictors and T-scores remained the same across all levels of age and education.

Reliable Change Classification

Linear regressions in the rNC sample were used to calculate three reliable change z-scores for the six neuropsychological measures: baseline performance predicting 12-month performance, 12-month performance predicting 24-month performance, and baseline performance predicting 24-month performance. Across these eighteen reliable change z-scores, means ranged from -0.003 to 0.003 and all standard deviations were 0.998 in the rNC sample. The percentage of rNC participants classified with a "significant increase" from time 1 to time 2 (e.g., z>1.645) across the eighteen z-scores ranged from 1.8% to 6.0%, the percentage with a "significant decrease" (z<-1.645) ranged from 3.5% to 7.4%, and the percentage with "stable" performance (-1.645≤ z ≤1.645) ranged from 88.0% to 93.7%. Chi-square tests comparing the proportion of rNCs with a significant increase or decrease for each of the 18 reliable change indices to the expected percentage under a 90% confidence interval (i.e., 5% at each tail) found no significant differences at α=0.01 (all *p*'s>0.039). Reliable change z-score formulas, adjusted R^2 for each model, and the proportion of rNC participants with a reliable decrease, increase, or stable performance is presented in Table 16.

Table 16: Regression-based reliable change z-score formulas derived from robust normal controls.

Time Points	Measure	Reliable Change Z-score Formula	Adjusted R^2	Proportion Decrease/Stable/Increase*
Baseline to 12-Months	Animal Fluency	[T-score.12 – (19.807 + (0.606 x T-score.0))]/7.933	.364	.049/.915/.035
	BNT	[T-score.12 – (20.847 + (0.586 x T-score.0))]/8.068	.341	.060/.898/.042
	TMT, Part A	[T-score.12 – (19.570 + (0.607 x T-score.0))]/7.944	.365	.035/.919/.046
	TMT, Part B	[T-score.12 – (23.619+ (0.530 x T-score.0))]/8.413	.280	.060/.880/.060
	AVLT Recall	[T-score.12 – (23.442 + (0.529 x T-score.0))]/8.508	.276	.056/.901/.042
	AVLT Recognition	[T-score.12 – (26.722 + (0.464 x T-score.0))]/8.837	.213	.046/.937/.018
12-Months to 24-Months	Animal Fluency	[T-score.24 – (17.768 + (0.645 x T-score.12))]/7.702	.408	.039/.915/.046
	BNT	[T-score.24 – (18.037 + (0.639 x T-score.12))]/7.612	.409	.060/.905/.035
	TMT, Part A	[T-score.24 – (17.677 + (0.646 x T-score.12))]/7.610	.416	.049/.095/.046
	TMT, Part B	[T-score.24 – (19.343 + (0.615 x T-score.12))]/7.817	.377	.060/.908/.032
	AVLT Recall	[T-score.24 – (16.833 + (0.664 x T-score.12))]/7.493	.439	.035/.908/.056
	AVLT Recognition	[T-score.24 – (25.229 + (0.494 x T-score.12))]/8.665	.242	.053/.915/.032
Baseline to 24-Months	Animal Fluency	[T-score.24 – (22.254 + (0.557 x T-score.0))]/8.361	.303	.042/.905/.053
	BNT	[T-score.24 – (21.066 + (0.581 x T-score.0))]/8.066	.337	.074/.898/.028
	TMT, Part A	[T-score.24 – (23.230 + (0.533 x T-score.0))]/8.436	.282	.049/.901/.049
	TMT, Part B	[T-score.24 – (24.306 + (0.517 x T-score.0))]/8.476	.268	.042/.919/.032
	AVLT Recall	[T-score.24 – (24.591 + (0.507 x T-score.0))]/8.639	.254	.049/.908/.042
	AVLT Recognition	[T-score.24 – (26.510 + (0.468 x T-score.0))]/8.809	.217	.063/.912/.025

Note: Robust Normal Control n=284. Reliable Decrease: $z<-1.645$; Stable: $z\geq-1.645$ and ≤1.645; Reliable Increase: $z>1.645$. All model *p's*≤0.001. BNT = 30-item Boston Naming Test; TMT = Trail Making Test; AVLT = Rey Auditory Verbal Learning Test

Serial Latent Profile Analyses

Two to eight latent class models were separately tested at baseline, 12-months, and 24-months using the multi-year demographically-corrected T-scores as indicators. Fit indices and descriptive characteristics for each model across the 3 time points are provided in Table 17. At all three time points the AIC and sBIC comparative fit indices continually decreased with increasing latent classes, and the BLRT showed a similar pattern such that *k* classes were always a statistically better fit than *k*-1 classes. These indices failed to clearly converge on an optimal solution, as this trend would presumably continue past eight latent classes and result in data over-fitting based on other indicators (Nylund et al., 2007).

The 6-class solution at baseline had the lowest BIC, though the VLMR-LRT suggested 6-classes would result in class over-extraction. The baseline VLMR-LRT indicated that 4-classes represented a better fit than 3-classes, and 5-classes provided a better fit than 4-classes. Entropy was almost equivalent with both the 4-class and 5-class solution, and neither model had a class size less than 5%. Given the combination of VLMR-LRT, satisfactory entropy, class size, model parsimony, and consistency with number of empirical MCI subtypes demonstrated in past research studies (Clark et al., 2013; Edmonds et al., 2014), the 4-class solution was chosen as the best fitting model at baseline.

At 12-months the solution with the lowest BIC was a 7-class solution, though both the VLMR-LRT and smallest class size (i.e., <5%) suggested a high likelihood of class over-extraction with this model. The VLMR-LRT indicated that the 4-class solution provided a better fit than the 3-class solution, the 5-class solution a better fit than the 4-class solution, and the 6-class solution a better fit than the 5-class solution. All of these solutions had smallest class sizes greater than 5%, though the 4-class solution had the highest entropy. Similar to the baseline LPA, the 4-class solution was chosen as the best fitting model at 12-months due to the combination of consistency with number of empirical MCI subtypes demonstrated in past research studies, highest entropy, VLMR-LRT, class size, and model parsimony.

The BIC index at 24-months demonstrated the same pattern of continuous decrease as the AIC, sBIC, and BLRT, suggesting likely class over-extraction. The VLMR-LRT suggested the 4-class model was a better fit than the 3-class model, though the 5-class model was not a better fit than the 4-class model. The 4-class solution also had the highest entropy of all solutions and its smallest class size was greater than 5% of the overall sample. Once again, the 4-class solution was chosen as the optimal model fit based on the VLMR-LRT, entropy, class size, model parsimony, and consistency with past research. Thus, a 4-class solution was selected as the best model fit at all three time points.

Table 17: LPA comparative fit indices, likelihood ratio tests and model characteristics in study 2.

Time Point	# of Classes	Final LL	AIC	BIC	sBIC	VLMR-LRT	BLRT	Entropy	Smallest Class Size
Baseline	2	-18558.522	37155.045	37244.637	37184.300	$p<0.0001$	$p<0.0001$	0.777	33%
	3	-18424.159	36900.317	37022.917	36940.351	$p=0.1232$	$p<0.0001$	0.697	26%
	4*	-18324.257	36714.515	36870.123	36765.327	$p<0.0001$	$p<0.0001$	0.749	16%
	5	-18270.770	36621.540	36810.156	36683.131	$p=0.0007$	$p<0.0001$	0.750	9%
	6	-18245.477	36584.955	36806.578	36657.323	$p=0.1410$	$p<0.0001$	0.754	7%
	7	-18223.805	36555.610	36810.241	36638.757	$p=0.0592$	$p<0.0001$	0.755	6%
	8	-18207.133	36536.265	36823.904	36630.190	$p=0.8440$	$p<0.0001$	0.708	6%
12-Months	2	-17025.204	34088.408	34176.215	34115.882	$p<0.0001$	$p<0.0001$	0.832	41%
	3	-16865.192	33782.384	33902.541	33819.981	$p=0.0003$	$p<0.0001$	0.769	30%
	4*	-16743.349	33552.699	33705.205	33600.417	$p<0.0001$	$p<0.0001$	0.801	12%
	5	-16690.190	33460.381	33645.237	33518.221	$p=0.0394$	$p<0.0001$	0.796	8%
	6	-16648.033	33390.066	33607.272	33458.029	$p=0.0061$	$p<0.0001$	0.792	8%
	7	-16621.628	33351.257	33600.813	33429.341	$p=0.3396$	$p<0.0001$	0.784	4%
	8	-16601.281	33324.563	33606.469	33412.770	$p=0.1694$	$p<0.0001$	0.792	4%
24-Months	2	-14450.324	28938.648	29023.386	28963.062	$p=0.0001$	$p<0.0001$	0.778	41%
	3	-14235.897	28523.794	28639.752	28557.204	$p<0.0001$	$p<0.0001$	0.818	18%
	4*	-14158.642	28383.283	28530.460	28425.687	$p=0.0123$	$p<0.0001$	0.820	8%
	5	-14107.167	28294.335	28472.731	28345.734	$p=0.1312$	$p<0.0001$	0.783	13%
	6	-14070.174	28234.348	28443.964	28294.742	$p=0.2304$	$p<0.0001$	0.805	6%
	7	-14044.275	28196.549	28437.384	28265.938	$p=0.2557$	$p<0.0001$	0.795	1%
	8	-14013.895	28149.790	28421.844	28228.173	$p=0.4432$	$p<0.0001$	0.810	5%

Note: *Chosen as best class solution. LPA = Latent Profile Analysis; LL= Log-Likelihood; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; sBIC = sample-size adjusted Bayesian Information Criterion; VLMR-LRT = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test; BLRT = Bootstrapped Likelihood Ratio Test

Neuropsychological Characterization of Latent Profile Classes

The final LPA solutions at baseline, 12-months, and 24-months classified participants into 4 analogous neuropsychological profiles at each time point: 1) a *multi-domain impairment class* (MLT) with deficits ranging from mild impairment to moderate impairment (Heaton et al., 2004) across all cognitive measures, 2) an *amnestic impairment class* (AMN) with deficits ranging from mild impairment to mild-to-moderate impairment only on memory measures, and below average to average performance on tests of language and attention/executive functions, 3) a *dysexecutive/below average cognition class* (DYS/BA) with scores in the mild impairment to below average range on tests of language and attention/executive functions, and below average scores on memory measures, and 4) an *average cognition class* (AVG)

with performance in the average range on tests of language and attention/executive functions, and in the average to above average range on memory measures despite their baseline MCI diagnosis. Final class counts and neuropsychological performance of each class at baseline, 12-, and 24-months is presented in Figure 10.

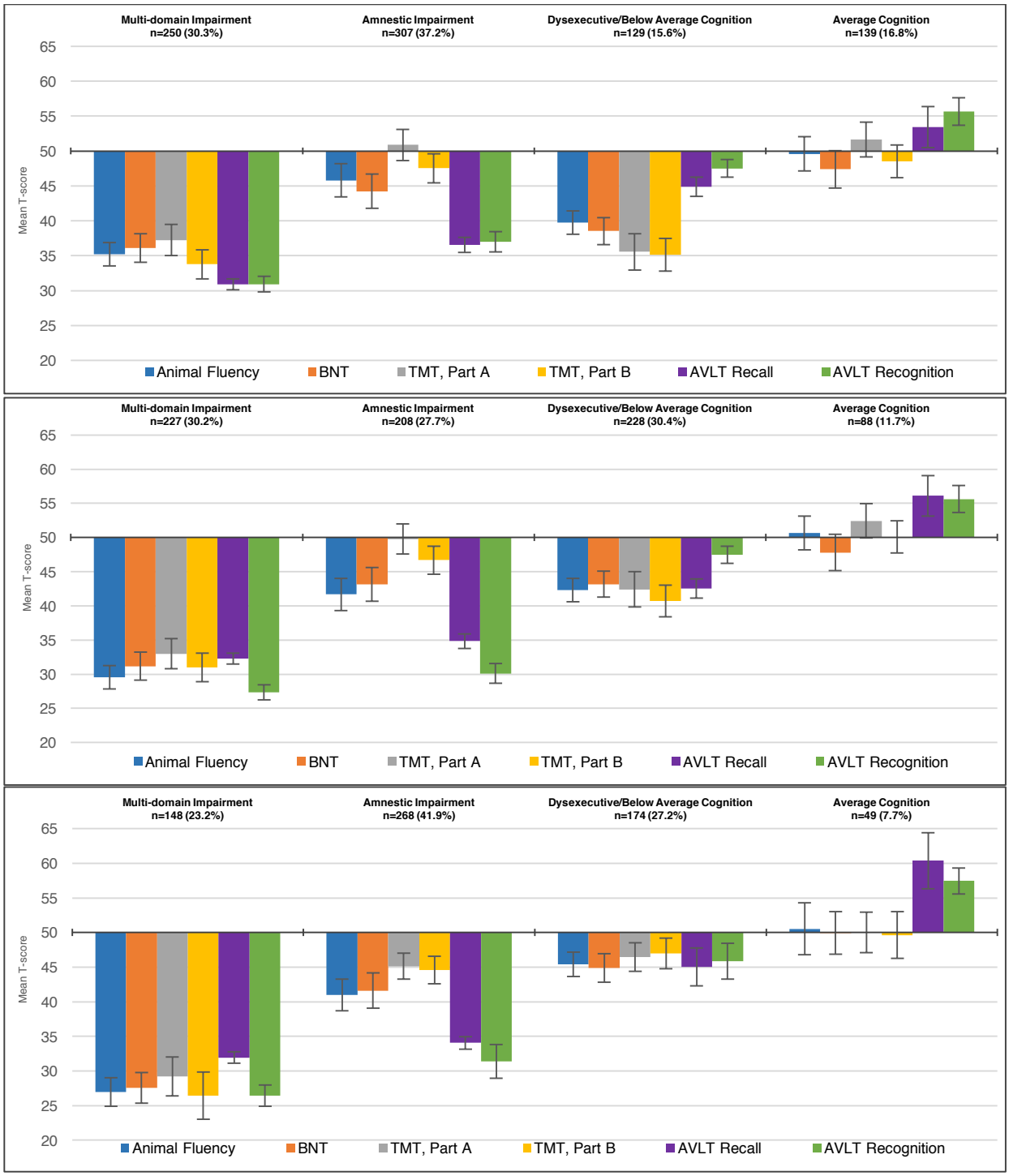

Figure 10: Neuropsychological profiles of the latent classes at baseline, 12-, and 24-months in study 2. Error bars denote 95% confidence intervals.
Top Panel: Neuropsychological performance at baseline (n=825).
Middle Panel: Neuropsychological performance at 12-months (n=751).
Bottom Panel: Neuropsychological performance at 24-months (n=639).
BNT = 30-item boston naming test; TMT = trail making test; AVLT = rey auditory verbal learning test.

Class sizes varied across time points, with the MLT class ranging from 23.2–30.3%, the AMN class from 27.7–41.9%, the DYS/BA class from 15.6–30.4%, and the AVG class from 7.7–16.8%. The AMN class was the largest class at baseline and 24-months, while the DYS/BA class was the largest size by a small margin at 12-months (the MLT class was almost equivalent in size at 12-months). The DYS/BA class was the smallest size at baseline and the AVG class the smallest at 12- and 24-months. Class size as a proportion of the total sample at each time point is presented in Figure 11.

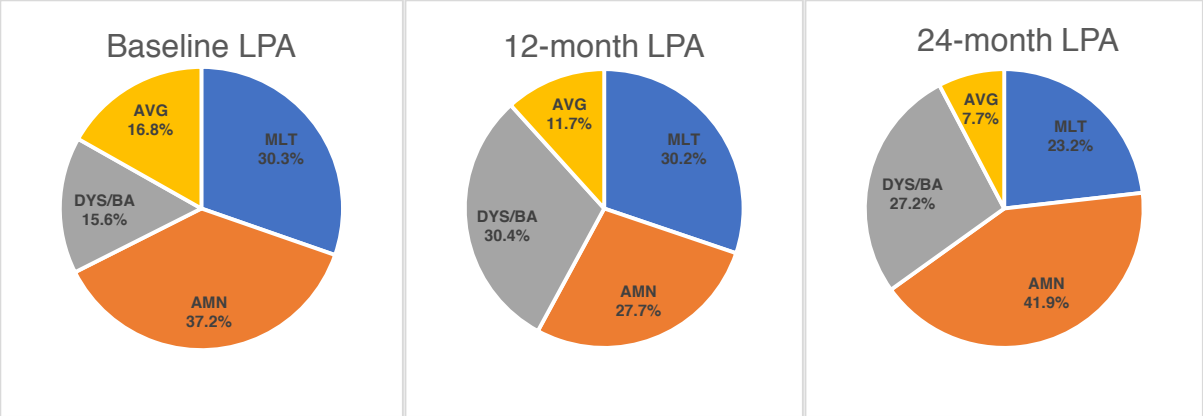

Figure 11: Difference in LPA class sizes at baseline, 12-, and 24-months in study 2.
Left Panel: Baseline LPA classes (n=825).
Middle Panel: 12-month LPA classes (n=751).
Right Panel: 24-month LPA classes (n=639).
LPA = latent profile analysis. MLT = multi-domain impairment; AMN = amnestic impairment; DYS/BA = dysexecutive/ below average cognition; AVG = average cognition.

Neuropsychological Parameter Comparisons

Omnibus Wald tests suggested differences between classes on every neuropsychological variable (all p's$\leq$0.0001) at each time point. Post-hoc comparisons with α= 0.008 (i.e., α= 0.05/6) indicated the MLT class performed worse than the AMN and AVG classes on all measures at each of the three time points (all p's$\leq$0.0035). The MLT class had worse performance on animal fluency and memory measures than the DYS/BA class at baseline (all p's$\leq$0.0036), with no differences in baseline performance on the BNT or attention/executive function measures (all p's$\geq$0.1822). At 12- and 24-months, the MLT class performed worse than the DYS/BA class on all measures (all p's$\leq$0.0001). The AMN class performed better than the DYS/BA class at baseline on measures of language and had better performance at baseline and 12-months on measures of attention/executive functions (all p's$\leq$0.001).

Across all three time points the AMN class demonstrated worse performance than the DYS/BA class on measures of memory, and at 24-months the AMN class also had worse animal fluency performance than the DYS/BA class (all *p's*≤0.0043). There were no differences between the AMN and DYS/BA classes on measures of language at 12-months, or on the BNT, TMT Part A, and TMT Part B at 24-months (all *p's*≥0.1571). The AVG class performed better than the AMN class on memory measures and animal fluency at all three time points and had better performance on the BNT and TMT Part A at 24-months (all *p's*≤0.0067). No differences between the AVG and AMN class were noted on the TMT Part B across all time points, or the BNT and TMT Part A at baseline and 12-months (all *p's*≥0.0102). The AVG class performed better than the DYS/BA class on all measures at baseline and 12-months, and at 24-months had better performance on the BNT and memory measures (all *p's*≤0.0074). There were no differences at 24-months between the AVG and DYS/BA classes on animal fluency or measures of attention/executive functions (all *p's*≥0.0124).

Omnibus Wald tests of within-class neuropsychological comparisons across the three time points indicated that the MLT class had worse performance over the 24-months on measures of language, attention/executive functions, and AVLT recognition (all *p's*≤0.0001). There was no difference in AVLT recall between the three time points (*p*=0.0821) for the MLT class. The AMN class performed worse over the 24-months on measures of memory, animal fluency, and TMT Part A (all *p's*≤0.0460), though no differences were noted on BNT or TMT Part B performances (all *p's*≥0.0841). The DYS/BA class demonstrated better performance over time on measures of language, attention/executive functions, and AVLT recall (all *p's*≤0.0371), with no difference in performance on AVLT recognition (*p*=0.5152). The AVG class had improved AVLT recall performance across the 24-months (*p*=0.0042) but did not otherwise demonstrate differences in the neuropsychological measures over time (all *p's*≥0.2070). Omnibus and post-hoc within-class differences in neuropsychological performance across the three time points is presented in Table 18.

Table 18: Within-class comparisons of neuropsychological performance for the 4 latent classes at baseline, 12-, and 24-months in study 2.

Latent Class	*Baseline* *(n= 825)*	*12-Months* *(n=751)*	*24-Months* *(n=639)*		
Multi-domain Impairment	**30.3% (n=250)**	**30.2% (n=227)**	**23.2% (n=148)**	**Omnibus Wald χ^2 (df)**	***p*-value**
Animal Fluency	35.240[2,3] (1.017)	29.554[1,3] (0.855)	26.968[1,2] (1.067)	$\chi^2(2)=43.510$	$p<0.0001$
BNT	36.107[2,3] (1.089)	31.174[1,3] (1.043)	27.571[1,2] (1.127)	$\chi^2(2)=38.655$	$p<0.0001$
TMT, Part A	37.246[2,3] (1.430)	33.004[1,3] (1.130)	29.213[1,2] (1.440)	$\chi^2(2)=20.102$	$p<0.0001$
TMT, Part B	33.779[3] (1.226)	31.004[3] (1.074)	26.435[1,2] (1.735)	$\chi^2(2)=15.270$	$p<0.0005$
AVLT Recall	30.892 (0.587)	32.280 (0.409)	31.929 (0.426)	$\chi^2(2)=4.999$	$p=0.0821$
AVLT Recognition	30.935[2,3] (0.586)	27.354[1] (0.575)	26.435[1] (0.772)	$\chi^2(2)=32.570$	$p<0.0001$
Amnestic Impairment	**37.2% (n=307)**	**27.7% (n=208)**	**41.9% (n=268)**	**Omnibus Wald χ^2 (df)**	***p*-value**
Animal Fluency	45.801[2,3] (0.972)	41.697[1] (1.200)	40.986[1] (1.152)	$\chi^2(2)=16.220$	$p=0.0003$
BNT	44.237 (0.981)	43.117 (1.258)	41.617 (1.307)	$\chi^2(2)=3.590$	$p=0.1662$
TMT, Part A	50.869[3] (0.880)	49.782[3] (1.135)	45.138[1,2] (0.944)	$\chi^2(2)=25.229$	$p<0.0001$
TMT, Part B	47.531 (1.087)	46.691 (1.052)	44.595 (1.023)	$\chi^2(2)=4.950$	$p=0.0841$
AVLT Recall	36.540[3] (0.926)	34.832 (0.557)	34.057[1] (0.471)	$\chi^2(2)=6.883$	$p=0.0320$
AVLT Recognition	37.002[2,3] (0.949)	30.115[1] (0.743)	31.395[1] (1.236)	$\chi^2(2)=6.358$	$p=0.0416$
Dysexecutive/Below Average Cognition	**15.6% (n=129)**	**30.4% (n=228)**	**27.2% (n=174)**	**Omnibus Wald χ^2 (df)**	***p*-value**
Animal Fluency	39.763[3] (1.075)	42.317[3] (0.865)	45.404[1,2] (0.883)	$\chi^2(2)=17.477$	$p=0.0002$
BNT	38.529[2,3] (1.326)	43.161[1] (0.971)	44.890[1] (1.040)	$\chi^2(2)=16.642$	$p<0.0001$
TMT, Part A	35.561[2,3] (1.185)	42.401[1,3] (1.323)	46.461[1,2] (1.067)	$\chi^2(2)=46.217$	$p<0.0001$
TMT, Part B	35.1282[2,3] (1.211)	40.718[1,3] (1.184)	47.008[1,2] (1.132)	$\chi^2(2)=52.626$	$p<0.0001$
AVLT Recall	44.879 (0.884)	42.530 (0.709)	45.065 (1.396)	$\chi^2(2)=6.586$	$p=0.0371$
AVLT Recognition	47.499 (1.537)	47.471 (0.640)	45.847 (1.334)	$\chi^2(2)=1.326$	$p=0.5152$

Table 18: Within-class comparisons of neuropsychological performance for the 4 latent classes at baseline, 12-, and 24-months in study 2, continued.

Latent Class	*Baseline* *(n= 825)*	*12-Months* *(n=751)*	*24-Months* *(n=639)*		
Average Cognition	**16.8% (n=139)**	**11.7% (n=88)**	**7.7% (n=49)**	**Omnibus Wald χ^2 (df)**	***p*-value**
Animal Fluency	49.589 (0.982)	50.638 (1.261)	50.522 (1.905)	$\chi^2(2)=0.702$	$p=0.7040$
BNT	47.380 (0.978)	47.785 (1.356)	49.964 (1.574)	$\chi^2(2)=2.529$	$p=0.2824$
TMT, Part A	51.639 (1.135)	52.445 (1.265)	50.012 (1.498)	$\chi^2(2)=2.432$	$p=0.2963$
TMT, Part B	48.528 (1.067)	50.079 (1.204)	49.636 (1.733)	$\chi^2(2)=1.495$	$p=0.4735$
AVLT Recall	53.414[3] (1.222)	56.115 (1.504)	60.363[1] (2.054)	$\chi^2(2)=10.945$	$p=0.0042$
AVLT Recognition	55.673 (0.953)	55.628 (1.007)	57.438 (0.966)	$\chi^2(2)=3.150$	$p=0.2070$

Note: Data summarized as mean (standard error) in T-scores. Numbered superscripts denote significant Wald χ^2 test post-hoc differences at $p \leq 0.0167$ ($\alpha=0.05/3$) within each analogous class across time points (1= Baseline, 2= 12-months, 3= 24-months). χ^2 = chi-square; df= degrees of freedom; BNT = 30-item boston naming test; TMT = trail making test; AVLT = rey auditory verbal learning test

Missing Data Analysis

Analysis of missing data indicated that there were no Omnibus Wald chi-square differences between classes at baseline or 12-months in the proportion of individuals that had missing data at follow-up time points (all *p*'s$\geq$0.140). Post-hoc differences between the classes at $\alpha=0.01$ for each time point were also non-significant (all *p*'s$\geq$0.035). Proportion of missing follow-up data by LPA class at baseline and 12-months is presented in Table 19.

Table 19: Missing participant follow-up data across latent profile classes at baseline and 12-months in study 2.

LPA Time Point	Missing Follow-up	Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition	Omnibus Wald χ^2 (df)	*p*-value
Baseline	12-months	8.6% (2.3)	7.3% (1.8)	11.0% (3.8)	8.4% (2.6)	$\chi^2(3)= 0.830$	$p=0.842$
Baseline	24-months	22.6% (3.2)	17.7% (2.6)	29.8% (4.9)	18.1% (3.6)	$\chi^2(3)= 5.482$	$p=0.140$
12-months	24-months	19.6% (2.9)	11.3% (2.7)	13.1% (2.7)	10.8% (3.5)	$\chi^2(3)= 5.279$	$p=0.152$

Note: Data summarized as mean (standard error) in percentages. There were no significant post-hoc differences at $\alpha=0.01$ between classes using the Wald χ^2 test (all *p*'s$\geq$0.035). LPA = latent profile analysis.

Additional analyses among the total LPA sample at each time point indicated a significant difference in baseline scores such that those with missing data at 12-months (n= 74) performed worse on baseline animal fluency (Mean T-score Difference=2.751, t[823]=2.163, *p*=0.031) than individuals with available data at 12-months (n=751), and participants with missing data at 24-months (n=186) performed worse on baseline animal fluency (Mean T-score Difference=2.034, t[823]=2.340, *p*=0.020), TMT Part A (Mean T-score Difference=2.555, t[823]=2.720, *p*=0.007), and TMT Part B (Mean T-score Difference=2.855, t[814]=3.038, *p*=0.002) than individuals with available data at 24-months (n=639). Further examination revealed a similar pattern in the overall sample at 12-months, such that participants with missing data at 24-months (n=112) had worse 12-month performance on the BNT (Mean T-score Difference=3.938, t(747)=2.948, *p*=0.003), Trails A (Mean T-score Difference=3.890, t[749]=3.187, *p*=0.001), and Trails B (Mean T-score Difference=3.478, t[744]=2.974, *p*=0.003) than individuals with available data at 24-months (n=639).

Demographic and Diagnostic Outcomes

Demographic differences including age, education and sex across time points, as well as differences in concurrent ADNI diagnosis at 12- and 24-months is presented in Table 20. There were differences in age, education, and sex between the baseline classes (all *p*'s$\leq$0.023). Post hoc comparisons at baseline indicated that the AVG class was younger than the AMN class, more educated than the DYS/BA class, and had a higher proportion of women than the AMN class (all *p*'s$\leq$0.008). The MLT class also had a greater percentage of women than the AMN and DYS/BA classes (all *p*'s$\leq$0.002). There were no differences in age, education, or sex between the classes at 12-months (all *p*'s$\geq$0.092). At 24-months, age and sex differed between classes (all *p*'s$\leq$0.007), though there was no difference in level of education (*p*=0.073). Post hoc comparisons at 24-months demonstrated that the AVG class was younger than the MLT, AMN, and DYS/BA classes, and the MLT class had a greater proportion of women than the DYS/BA class (all *p*'s$\leq$0.007).

Examination of concurrent ADNI diagnosis at 12- and 24-months (all participants were diagnosed by ADNI with MCI at baseline) found diagnostic differences between classes at the time points (all

p's≤0.001). Post-hoc comparisons indicated that at 12- and 24-months, the MLT class had a higher percentage of participants diagnosed by ADNI with dementia and a lower proportion with MCI than the AMN, DYS/BA, and AVG classes (all *p's*<0.001). The AVG class at 12- and 24-months had a higher percentage of individuals that were identified by ADNI as normal and a lower proportion diagnosed with dementia than the AMN class (all *p's*≤0.004). The DYS/BA class at 24-months also had a greater proportion of participants identified as normal and a lower percentage diagnosed with dementia than the AMN class (*p*<0.001).

Table 20: Demographic differences between latent classes at baseline, 12-, and 24-months in study 2.

Time Point	Demo. Variable	Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition	Omnibus Wald χ^2 (df)	*p*-value
Baseline	Age	73.061 (0.491)	74.937[4] (0.460)	74.113 (0.776)	72.619[2] (0.704)	$\chi^2(3)$= 9.518	*p*=0.023
	Education	15.964 (0.213)	15.916 (0.189)	15.034[4] (0.334)	16.642[3] (0.234)	$\chi^2(3)$= 14.610	*p*=0.002
	Sex (%)	55.9% F[2,3] (4.6)	28.3% F[1,4] (3.3)	33.5% F[1] (5.1)	47.1% F[2] (4.7)	$\chi^2(3)$= 24.844	*p*<0.001
12-Months	Age	74.256 (0.483)	75.741 (0.571)	75.116 (0.530)	73.318 (0.900)	$\chi^2(3)$= 6.429	*p*=0.092
	Education	16.036 (0.227)	15.682 (0.211)	15.979 (0.220)	16.512 (0.293)	$\chi^2(3)$= 5.348	*p*=0.148
	Sex (%)	46.5% F (4.0)	33.2% F (4.5)	38.4% F (3.8)	45.2% F (5.8)	$\chi^2(3)$= 4.811	*p*=0.186
	12-Months Diagnosis (%)	NL: 0.0% (0.0)[2,3,4] MCI: 66.2% (3.9) DM: 33.8% (3.9)	NL: 2.4% (1.3)[1,4] MCI: 91.1% (2.4) DM: 6.5% (2.3)	NL: 3.8% (1.6)[1] MCI: 92.5% (2.0) DM: 3.7% (1.6)	NL: 9.0% (3.1)[1,2] MCI: 91.0% (3.1) DM: 0.0% (0.0)	$\chi^2(6)$= 127.437	*p*<0.001
24-Months	Age	74.811[4] (0.567)	76.425[4] (0.458)	76.097[4] (0.663)	71.473[1,2,3] (1.001)	$\chi^2(3)$= 21.969	*p*<0.001
	Education	16.096 (0.263)	15.926 (0.190)	15.867 (0.244)	16.883 (0.337)	$\chi^2(3)$= 6.981	*p*=0.073
	Sex (% F)	52.1% F[3] (4.6)	36.8% F (4.2)	31.4% F[1] (5.7)	50.9% F (8.0)	$\chi^2(3)$= 12.006	*p*=0.007
	24-Months Diagnosis (%)	NL: 0.0% (0.0)[2,3,4] MCI: 27.6% (4.8) DM: 72.4% (4.8)	NL: 1.6% (0.9)[1,3,4] MCI: 79.6% (3.3) DM: 18.8% (2.3)	NL: 12.0% (2.9)[1,2] MCI: 86.0% (3.0) DM: 2.0% (1.7)	NL: 17.6% (5.8)[1,2] MCI: 82.4% (5.8) DM: 0.0% (0.0)	$\chi^2(6)$= 297.130	*p*<0.001

Note: Data summarized as mean (standard error) in T-scores. Numbered superscripts denote significant Wald χ^2 test post-hoc differences at *p*<0.0083 (0.05/6) between classes (1= Multi-domain impairment, 2= Amnestic Impairment, 3= Dysexecutive/Below Average Cognition, 4= Average Cognition. Demo.= Demographic; χ^2 = chi-square; df= degrees of freedom; F = Female; NL = Normal; MCI = Mild Cognitive Impairment DM = Dementia

Reliable Change Outcomes

Reliable change classification between classes at 12- and 24-months is presented in Figure 12.

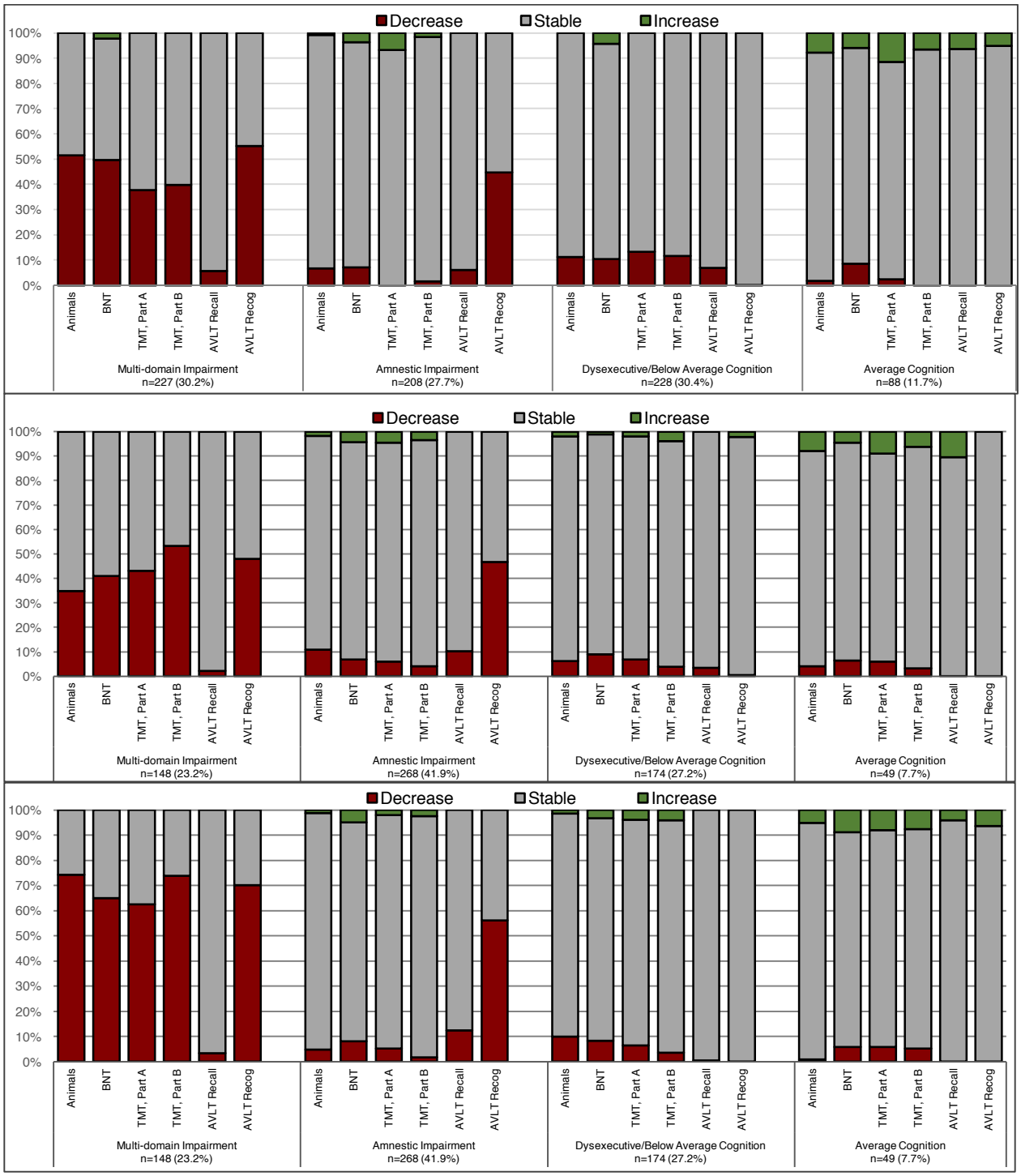

Figure 12: Reliable change within latent classes at 12- and 24-months in study 2.
Top Panel: Reliable change from baseline to 12-months in the 12-month LPA (n=751).
Middle Panel: Reliable change from 12- to 24-months in the 24-month LPA (n=639).
Bottom Panel: Reliable change from baseline to 24-months in the 24-month LPA (n=639).
Decrease: z< -1.645, Stable: -1.645≤ z ≤1.645, Increase: z> 1.645. Animals = Animal Fluency; BNT = 30-item Boston Naming Test; TMT = Trail Making Test; AVLT = Rey Auditory Verbal Learning Test; Recog = Recognition.

Omnibus Wald tests between classes at 12-months and 24-months found differences in reliable change classification (i.e., significant decrease, stable, significant increase) for all neuropsychological measures (all *p*'s<0.001). Post hoc comparisons between the 12-month LPA classes indicated that the MLT class had a greater proportion of participants with a significant decrease in baseline to 12-month performance on measures of language and attention/executive functions than the AMN, DYS/BA, and AVG classes (all *p*'s<0.008). Both the MLT and AMN classes had a higher percentage of individuals with a significant decrease in baseline to 12-month performance on AVLT recall than the AVG class, and on AVLT recognition than the DYS/BA and AVG classes (all *p*'s<0.001), though the MLT and AMN classes did not differ from each other on memory measures. The DYS/BA class at 12-months had a greater proportion of participants with a significant decrease in baseline to 12-month performance on measures of attention/executive functions than the AMN and AVG classes, and a higher percentage with a significant decrease on AVLT recall than the AVG class (all *p*'s<0.008).

Post hoc comparisons between the 24-month LPA classes indicated that the MLT class had a greater proportion of participants with a significant decrease in 12- to 24-month and baseline to 24-month performance on measures of language and attention/executive functions than the AMN, DYS/BA, and AVG classes (all *p*'s<0.001). The MLT class also had a higher percentage of individuals with a significant decrease in 12- to 24-month and baseline to 24-month performance on AVLT recognition than the DYS/BA and AVG classes (all *p*'s<0.001), though no differences were found compared to the AMN class. The AMN class had a greater proportion of individuals with a significant decrease from 12- to 24-month AVLT recall performance than the AVG class, as well as a larger percentage of participants with a significant decrease in baseline to 24-month AVLT recall performance than the MLT, DYS/BA, and AVG classes (all *p*'s<0.008). There were no differences between the DYS/BA and AVG classes in reliable change classification from 12- to 24-month or baseline to 24-month performance on any of the neuropsychological measures. Differences in reliable change classification for the 12- and 24-month LPA classes are presented in Table 21.

Table 21: Differences in reliable classification between 12- and 24-month latent classes in study 2.

Time Point	Measure	Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition	Omnibus Wald χ^2 (df)	*p*-value
Baseline to 12-Months	Animal Fluency	**–:** 51.6% (5.3)[2,3,4] **=:** 48.4% (5.3) **+:** 0.0% (0.0)	**–:** 6.8% (3.3)[1] **=:** 92.4% (3.3) **+:** 0.9% (0.7)	**–:** 11.3% (2.6)[1] **=:** 88.7% (2.6) **+:** 0.0% (0.0)	**–:** 1.7% (2.8)[1] **=:** 90.4% (3.5) **+:** 7.9% (3.3)[2,3]	$\chi^2(6)=$ 68.580	*p*<0.001
	BNT	**–:** 49.6% (4.7)[2,3,4] **=:** 48.1% (4.5) **+:** 2.3% (1.3)	**–:** 7.1% (3.1)[1] **=:** 89.2% (3.0) **+:** 3.7% (1.6)	**–:** 10.5% (2.6)[1] **=:** 85.2% (2.9) **+:** 4.4% (1.5)	**–:** 8.5% (3.6)[1] **=:** 85.6% (4.2) **+:** 5.9% (2.7)	$\chi^2(6)=$ 63.607	*p*<0.001
	TMT, Part A	**–:** 37.8% (5.3)[2,3,4] **=:** 62.2% (4.3) **+:** 0.0% (0.0)	**–:** 0.0% (0.0)[1,3] **=:** 93.2% (1.9) **+:** 6.8% (1.9)	**–:** 13.2% (3.9)[1,2,4] **=:** 86.8% (3.9) **+:** 0.0% (0.0)	**–:** 2.3% (2.3)[1,3] **=:** 86.3% (4.3) **+:** 11.5% (3.8)	$\chi^2(6)=$ 137.835	*p*<0.001
	TMT, Part B	**–:** 39.8% (7.3)[2,3,4] **=:** 60.2% (7.3) **+:** 0.0% (0.0)	**–:** 1.6% (1.4)[1,3] **=:** 96.8% (1.5) **+:** 1.7% (0.9)	**–:** 11.6% (3.3)[1,2,4] **=:** 88.4% (3.3) **+:** 0.0% (0.0)	**–:** 0.0% (0.0)[1,3] **=:** 93.5% (2.8) **+:** 6.5% (2.8)	$\chi^2(6)=$ 86.548	*p*<0.001
	AVLT Recall	**–:** 5.6% (1.7)[4] **=:** 94.4% (1.7) **+:** 0.0% (0.0)	**–:** 6.1% (2.0)[4] **=:** 93.9% (2.0) **+:** 0.0% (0.0)	**–:** 7.0% (2.0)[4] **=:** 93.0% (2.0) **+:** 0.0% (0.0)	**–:** 0.0% (0.0)[1,2,3] **=:** 93.6% (2.7) **+:** 6.4% (2.7)	$\chi^2(6)=$ 50.096	*p*<0.001
	AVLT Recognition	**–:** 55.1% (4.5)[3,4] **=:** 44.9% (4.5) **+:** 0.0% (0.0)	**–:** 44.8% (5.7)[3,4] **=:** 55.2% (5.7) **+:** 0.0% (0.0)	**–:** 0.2% (0.5)[1,2] **=:** 99.8% (0.5) **+:** 0.0% (0.0)	**–:** 0.0% (0.0)[1,2] **=:** 94.8% (2.3) **+:** 5.2% (2.3)	$\chi^2(6)=$ 344.729	*p*<0.001

Time Point	Measure	Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition	Omnibus Wald χ^2 (df)	*p*-value
12-Months to 24-Months	Animal Fluency	**–:** 34.9% (4.5)[2,3,4] **=:** 65.1% (4.5) **+:** 0.0% (0.0)	**–:** 10.8% (4.3)[1] **=:** 87.4% (4.1) **+:** 1.8% (1.0)	**–:** 6.3% (2.2)[1] **=:** 91.7% (2.4) **+:** 2.0% (1.3)	**–:** 4.2% (3.0)[1] **=:** 87.8% (5.6) **+:** 8.0% (4.7)	$\chi^2(6)=$ 51.639	*p*<0.001
	BNT	**–:** 41.0% (4.4)[2,3,4] **=:** 59.0% (4.4) **+:** 0.0% (0.0)	**–:** 6.8% (5.9)[1] **=:** 88.8% (5.3) **+:** 4.4% (1.6)	**–:** 8.9% (3.3)[1] **=:** 90.0% (3.0) **+:** 1.1% (1.4)	**–:** 6.4% (3.7)[1] **=:** 89.1% (4.6) **+:** 4.5% (3.4)	$\chi^2(6)=$ 64.824	*p*<0.001
	TMT, Part A	**–:** 43.0% (5.8)[2,3,4] **=:** 57.0% (5.8) **+:** 0.0% (0.0)	**–:** 6.1% (2.2)[1] **=:** 89.6% (2.3) **+:** 4.5% (1.4)	**–:** 6.9% (2.7)[1] **=:** 91.2% (2.7) **+:** 1.9% (1.4)	**–:** 5.9% (3.4)[1] **=:** 85.2% (5.1) **+:** 8.9% (4.1)	$\chi^2(6)=$ 62.186	*p*<0.001
	TMT, Part B	**–:** 53.2% (7.0)[2,3,4] **=:** 46.8% (7.0) **+:** 0.0% (0.0)	**–:** 4.2% (1.7)[1] **=:** 92.4% (2.0) **+:** 3.4% (1.3)	**–:** 3.9% (3.8)[1] **=:** 92.3% (3.6) **+:** 3.8% (2.0)	**–:** 3.2% (3.2)[1] **=:** 90.6% (4.5) **+:** 6.2% (3.8)	$\chi^2(6)=$ 79.198	*p*<0.001
	AVLT Recall	**–:** 2.2% (1.3) **=:** 97.8% (1.3) **+:** 0.0% (0.0)	**–:** 10.2% (2.2)[4] **=:** 89.8% (2.2) **+:** 0.0% (0.0)	**–:** 3.5% (2.0) **=:** 96.5% (2.0) **+:** 0.0% (0.0)	**–:** 0.0% (0.0)[2] **=:** 87.9% (4.7) **+:** 10.3% (4.7)	$\chi^2(6)=$ 43.121	*p*<0.001
	AVLT Recognition	**–:** 48.0% (7.6)[3,4] **=:** 52.0% (7.6) **+:** 0.0% (0.0)	**–:** 46.8% (8.0)[3,4] **=:** 53.2% (8.0) **+:** 0.0% (0.0)	**–:** 0.4% (1.2)[1,2] **=:** 97.4% (1.5) **+:** 2.1% (1.1)	**–:** 0.0% (0.0)[1,2] **=:** 100.0% (0.0) **+:** 0.0% (0.0)	$\chi^2(6)=$ 260.137	*p*<0.001

Table 21: Differences in reliable classification between 12- and 24-month latent classes in study 2, continued.

Time Point	Measure	Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition	Omnibus Wald χ^2 (df)	*p*-value
Baseline to 24-Months	Animal Fluency	**–:** 74.2% (5.5)[2,3,4] **=:** 25.8% (5.5) **+:** 0.0% (0.0)	**–:** 4.7% (6.2)[1] **=:** 94.1% (6.1) **+:** 1.2% (0.7)	**–:** 9.9% (3.4)[1] **=:** 88.7% (3.4) **+:** 1.5% (1.1)	**–:** 0.8% (5.6)[1] **=:** 94.0% (5.2) **+:** 5.2% (3.9)	χ^2(6)= 172.454	*p*<0.001
	BNT	**–:** 65.0% (5.7)[2,3,4] **=:** 35.0% (5.7) **+:** 0.0% (0.0)	**–:** 8.1% (5.4)[1] **=:** 87.0% (5.0) **+:** 4.9% (1.6)	**–:** 8.3% (2.8)[1] **=:** 88.5% (3.0) **+:** 3.2% (1.6)	**–:** 5.8% (3.5)[1] **=:** 85.4% (5.4) **+:** 8.8% (4.3)	χ^2(6)= 120.069	*p*<0.001
	TMT, Part A	**–:** 62.4% (8.5)[2,3,4] **=:** 37.6% (8.5) **+:** 0.0% (0.0)	**–:** 5.3% (2.1)[1] **=:** 92.6% (2.0) **+:** 2.1% (1.3)	**–:** 6.5% (3.2)[1] **=:** 89.6% (2.9) **+:** 3.9% (2.2)	**–:** 5.9% (3.5)[1] **=:** 86.0% (5.3) **+:** 8.1% (4.3)	χ^2(6)= 86.195	*p*<0.001
	TMT, Part B	**–:** 73.9% (5.9)[2,3,4] **=:** 26.1% (5.9) **+:** 0.0% (0.0)	**–:** 1.7% (1.1)[1] **=:** 95.9% (1.4) **+:** 2.4% (1.0)	**–:** 3.5% (2.8)[1] **=:** 92.4% (3.2) **+:** 4.1% (2.0)	**–:** 5.3% (5.3)[1] **=:** 87.1% (6.0) **+:** 7.6% (4.2)	χ^2(6)= 182.446	*p*<0.001
	AVLT Recall	**–:** 3.3% (1.8)[2] **=:** 96.7% (1.8) **+:** 0.0% (0.0)	**–:** 12.5% (2.2)[1,3,4] **=:** 87.5% (2.2) **+:** 0.0% (0.0)	**–:** 0.4% (0.7)[2] **=:** 99.6% (0.7) **+:** 0.0% (0.0)	**–:** 0.0% (0.0)[2] **=:** 95.9% (2.9) **+:** 4.1% (2.9)	χ^2(6)= 42.594	*p*<0.001
	AVLT Recognition	**–:** 70.2% (9.0)[3,4] **=:** 29.8% (9.0) **+:** 0.0% (0.0)	**–:** 56.1% (14.2)[3,4] **=:** 43.9% (14.2) **+:** 0.0% (0.0)	**–:** 0.1% (1.2)[1,2] **=:** 99.9% (1.2) **+:** 0.0% (0.0)	**–:** 0.0% (0.0)[1,2] **=:** 93.7% (2.8) **+:** 6.3% (2.8)	χ^2(6)= 385.917	*p*<0.001

Note: Data summarized as mean (standard error) in T-scores. Numbered superscripts denote significant Wald χ^2 test post-hoc differences at $p<0.0083$ (0.05/6) between classes (1= Multi-domain impairment, 2= Amnestic impairment, 3= Dysexecutive/Below average cognition, 4= Average cognition). χ^2 = chi-square; df= degrees of freedom; "-" = significant decrease; "=" = stable; "+" = significant increase; BNT = 30-item boston naming Test; TMT = trail making test; AVLT = rey auditory verbal learning test

Study 3: A latent transition model examining the likelihood of class change in neuropsychological subtypes of mild cognitive impairment.

Measurement Invariance

LPA models of measurement invariance were established by specifying four classes for each time point and using starting values from the LPA means in study 2 as class indicators. The Satorra–Bentler scaled chi-square difference test indicated that an LPA model with full measurement invariance (i.e., all means constrained to be equal within each analogous class across the three time points) had worse fit the base model with all parameter means freely estimated at baseline, 12-months, and 24-months (Satorra–Bentler $\chi^2[48]=323.411$, $p<0.0001$). Partial measurement invariance was examined by constraining means across baseline, 12-, and 24-months only for the neuropsychological indicators that

did not demonstrate significant differences within each analogous class over the three time points. Given the omnibus Wald test results of within-class comparisons in study 2, AVLT recall was constrained in the MLT class, BNT and TMT Part B were constrained in the AMN class, AVLT recognition was constrained in the DYS/BA class, and animal fluency, BNT, TMT Part A, TMT Part B, and AVLT recognition were constrained in the AVG class. Parameter means for all other neuropsychological variables within each class were freely estimated across time points. The Satorra–Bentler scaled chi-square difference test indicated no statistical difference in fit between this model of partial measurement invariance and the base model with all parameter means freely estimated at baseline, 12-months, and 24-months (Satorra–Bentler $\chi^2[18]=27.277$, $p=0.0739$). All additional analyses used this model of partial measurement invariance and constrained the aforementioned parameter means to be equal over time within each respective class.

Classes in Latent Transition Analysis

A model of latent transition analysis was developed by specifying four classes for each time point, using starting values derived from the LPA means in study 2 as class indicators, and constraining the appropriate parameters over time for each class to establish partial measurement invariance. MPlus uses Bayesian analysis to perform multiple imputation of missing data under the missing at random assumption, allowing for LTA analyses to use the neuropsychological performance of all 825 participants as class indicators at each time point. Similar to study 2, the latent transition analysis produced a *multi-domain impairment class* (MLT) with deficits ranging from mild impairment to moderate impairment across all cognitive measures, 2) an *amnestic impairment class* (AMN) with deficits ranging from mild impairment to mild-to-moderate impairment only on memory measures, and below average to average performance on tests of language and attention/executive functions, 3) a *dysexecutive/below average cognition class* (DYS/BA) with attention/executive function test scores in the mild impairment range, language test scores in the mild impairment to below average range, and memory test scores in the below average to average range, and 4) an *average cognition class* (AVG) with performance in the average range across all cognitive measures. Unlike its counterpart class in the separate LPAs of study 2, the DYS/BA class demonstrated relatively equivalent performance within a cognitive domain across all three time points

(e.g., mild attention/executive function impairment at baseline, 12-, and 24-months). Additionally, the AVG class did not demonstrate better memory performance over time, with scores remaining in the average range for both AVLT recall and recognition at baseline, 12-, and 24-months. Class size was also very consistent within analogous classes over the 24-months, further indicating that the LTA classes were highly stable across the three time points. High entropy (0.906) suggested that the LTA model fit the data well. Neuropsychological performance within each LTA class at baseline, 12-, and 24-months is presented in Figure 13.

Figure 13: Classes in the latent transition model in study 3. Error bars denote 95% confidence intervals.
1st Panel: Multi-domain impairment class. *2nd Panel:* Dysexecutive/below average cognition class.
3rd Panel: Amnestic impairment class. *4th Panel:* Average cognition class.
BNT = 30-item boston naming test; TMT = trail making test; AVLT = rey auditory verbal learning test.

Latent Transition Probabilities

Latent transition probabilities are presented in Table 22. Across all classes participants had a high likelihood of remaining within their analogous class (86.7–98.6%) from baseline to 12-months, and 12-months to 24-months, and a relatively small probability of transition to another class at the subsequent time point (0.7–8.3%). The MLT class had the least likelihood of transition such that 98–99% remained in the MLT class over time. Participants in the AMN class also had a high probability of staying in the AMN class (95–96%) over time, albeit with a very small but consistent likelihood (2–3%) of transition to the AVG class from baseline to 12-months and 12-months to 24-months. DYS/BA participants were slightly more likely to remain within the DYS/BA class from baseline to 12-months (92%) than 12-months to 24-months (87%), and they demonstrated variable, albeit small probabilities of transition to each of the other three classes over the three time points (1–7%). Individuals in the AVG class had a slightly lower likelihood to remain within the AVG class from baseline to 12-months (87%) than 12-months to 24-months (93%), and they demonstrated a modest but consistent probability of transition to the AMN class (7–8%). A graphical representation of the model of latent transition probabilities is presented in Figure 14.

Table 22: Latent transition probabilities between classes from baseline to 12-months and 12-months to 24-months.

		12-months			
		Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition
Baseline	Multi-domain Impairment	0.986	0.000	0.014	0.000
	Amnestic Impairment	0.010	0.960	0.000	0.031
	Dysexecutive/ Below Average Cognition	0.057	0.000	0.919	0.023
	Average Cognition	0.000	0.083	0.050	0.867
		24-months			
		Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition
12-months	Multi-domain Impairment	0.980	0.000	0.020	0.000
	Amnestic Impairment	0.000	0.956	0.018	0.026
	Dysexecutive/ Below Average Cognition	0.007	0.056	0.870	0.068
	Average Cognition	0.000	0.068	0.000	0.932

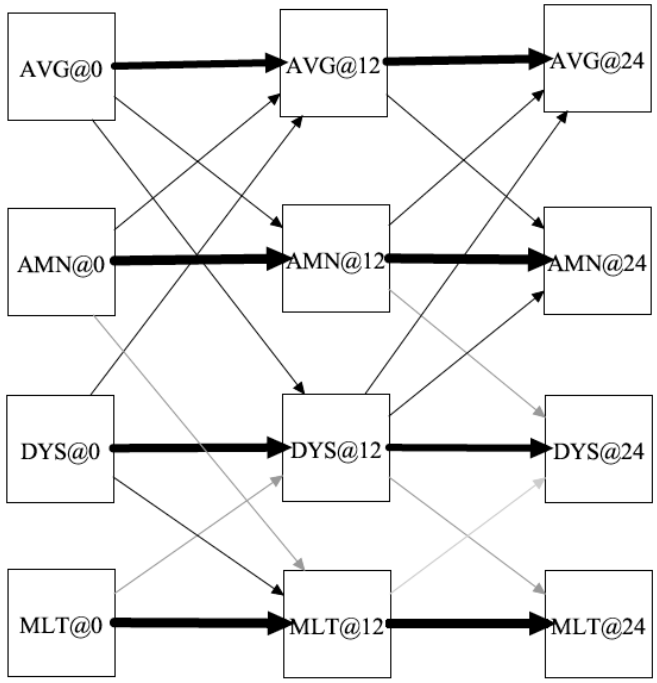

Figure 14: Graphical representation of the model of latent transition probabilities. Increasing arrow darkness and thickness indicates a higher likelihood of transition between classes across time. AVG = average cognition class; AMN = amnestic impairment class; DYS = dysexecutive/below average cognition class; MLT = multi-domain impairment class; @0 = at baseline; @12 = at 12-months; @24 = at 24-months.

Model inclusion of a second-order effect investigating the lasting direct effect that class membership at baseline had on class membership at 24-months produced little change in transition probabilities. Participants across all classes continued to have a high likelihood of remaining within their analogous class (>85%) from baseline to 12-months, and 12-months to 24-months, with a relatively small probability of transition to another class at the subsequent time point (<9%). Individuals in the MLT and AMN classes continued to have the highest likelihoods of remaining within their class over the three time points, (MLT: 98–99%, AMN: 94-96%). There was also a small but consistent probability (3–4%) for AMN class participants to transition to the AVG class from baseline to 12-months and 12-months to 24-months, although they also demonstrated a very low likelihood of transition to the DYS/BA class across the three occasions with the second-order effect (1-2%). A similar pattern of remaining within class from baseline to 12-months (92%) and 12-months to 24-months (85%) was observed for the DYS/BA class, again with small, variable probabilities of transition to each of the other three classes over the three time points (1–7%). There was also a similar likelihood to remain within class for AVG participants from baseline to 12-

months (87%) and 12-months to 24-months (92%), as well as a consistent probability to transition to the AMN class (7–8%). Latent transition probabilities in a model with a second-order effect are presented in Table 23.

Table 23: Latent transition probabilities between classes from baseline to 12-months and 12-months to 24-months with inclusion of a direct second-order effect of baseline to 24-months.

		12-months			
		Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition
Baseline	Multi-domain Impairment	0.987	0.000	0.013	0.000
	Amnestic Impairment	0.010	0.937	0.013	0.040
	Dysexecutive/ Below Average Cognition	0.058	0.000	0.920	0.022
	Average Cognition	0.000	0.084	0.050	0.865
		24-months			
		Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition
12-months	Multi-domain Impairment	0.978	0.000	0.022	0.000
	Amnestic Impairment	0.000	0.957	0.014	0.029
	Dysexecutive/ Below Average Cognition	0.010	0.068	0.854	0.067
	Average Cognition	0.000	0.080	0.000	0.920

Effect of Covariates on Latent Transition Probabilities

Available covariate data by class in the latent transition model is presented in Table 24. More than 99% of participants across all classes had available FAQ and APOE genotype information. AD-CSF biomarkers and ATN classification availability ranged from 65.3 – 75.1% across the LTA classes.

Table 24: Available covariate data by class in the latent transition model (n=825) of study 3.

Covariate	Multi-domain Impairment	Amnestic Impairment	Dysexecutive/ Below Average Cognition	Average Cognition
FAQ (n=822)	99.1%	100%	99.4%	100%
APOE genotype (n=821)	99.1%	99.6%	100%	99.3%
AD-CSF biomarkers (n= 591)	68.0% – 68.6%	71.3% – 72.4%	74.4% – 75.1%	72.1% – 72.8%
ATN classification (n=566)	67.1% – 67.7%	68.1% – 69.9%	71.0% – 71.8%	65.3% –67.1%

Note: FAQ = functional activities questionnaire, APOE = apolipoprotein E; AD = Alzheimer's disease; CSF = cerebrospinal fluid; ATN = Amyloid/Tau/Neurodegeneration.

The effect of FAQ on transition probabilities is presented in Table 25. MLT participants had a very high probability to remain within class across the three time points regardless of FAQ score (97–100%). Individuals in the AMN class were very likely to remain within class across the three time points when FAQ<6 (95–96%), but had a lower probability to stay within class from baseline to 12-months when FAQ≥6 (84%) as well as a greater likelihood to transition to the MLT class (11%). DYS/BA class participants were most likely to remain within class across the three time points when FAQ<6 (92–94%). However, individuals in the DYS/BA class had a lower probability to remain within class when FAQ≥6 (70-86%) and also demonstrated a modest likelihood of transition to the MLT class from baseline to 12-months (11%) and a moderate likelihood of transition to the AMN class from 12- to 24-months (30%). AVG class participants had a high probability to remain within class when FAQ<6 across the three time points (84-85%) as well as small likelihoods of transition to the DYS/BA (5-7%) and AMN (7-9%) classes. FAQ≥6 had a variable effect on the likelihood of individuals in the AVG class staying within class (80–96%), though it increased the transition probability of AVG participants to the DYS/BA class from baseline to 12-months (20%).

Table 25: Latent transition probabilities between classes from baseline to 12-months and 12- to 24-months by baseline FAQ score.

		FAQ <6						**FAQ ≥6**			
		12-months						*12-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
Baseline	MLT	0.973	0.000	0.027	0.000	*Baseline*	MLT	1.000	0.000	0.000	0.000
	AMN	0.000	0.964	0.000	0.036		AMN	0.111	0.839	0.050	0.000
	DYS/BA	0.050	0.000	0.921	0.029		DYS/BA	0.142	0.000	0.858	0.000
	AVG	0.000	0.099	0.054	0.847		AVG	0.000	0.000	0.203	0.797
		24-months						*24-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
12-months	MLT	0.969	0.000	0.000	0.031	*12-months*	MLT	0.997	0.000	0.000	0.003
	AMN	0.000	0.947	0.034	0.019		AMN	0.000	1.000	0.000	0.000
	DYS/BA	0.000	0.064	0.936	0.000		DYS/BA	0.000	0.301	0.699	0.000
	AVG	0.013	0.072	0.072	0.843		AVG	0.000	0.000	0.045	0.955

Note: FAQ = functional activities questionnaire; MLT = multi-domain impairment; AMN = amnestic impairment; DYS/BA = dysexecutive/ below average cognition; AVG = average cognition.

The effect of APOE ε4 allele presence on transition probabilities is presented in Table 26. MLT participants had a very high probability to remain within class across the three time points regardless of APOE ε4 allele presence (94–100%); a similar pattern was observed regarding the effect of APOE ε4 allele presence on the transition probabilities of individuals in the AMN class (94–98%). DYS/BA class participants were most likely to remain within class across the three time points with APOE ε4-negativity (95%). However, individuals in the DYS/BA class had a lower probability to remain within class in the presence of APOE ε4-positivity (75–90%) as well as a modest likelihood of transition to the AMN class from 12- to 24-months (14%). AVG class participants had a high probability to remain within class across the three time points regardless of APOE ε4 allele presence (87–94%), though the likelihood of transition to the AMN class demonstrated a very small increase with APOE ε4-positivity (6–8%) compared to APOE ε4-negativity (8–9%).

Table 26: Latent transition probabilities between classes from baseline to 12-months and 12- to 24-months by APOE ε4 allele presence.

		APOE ε4 negative						**APOE ε4 positive**			
		12-months						*12-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
Baseline	MLT	0.936	0.000	0.064	0.000	*Baseline*	MLT	1.000	0.000	0.000	0.000
	AMN	0.000	0.943	0.011	0.046		AMN	0.028	0.967	0.000	0.005
	DYS/BA	0.054	0.000	0.946	0.000		DYS/BA	0.045	0.000	0.897	0.057
	AVG	0.000	0.076	0.057	0.867		AVG	0.000	0.091	0.000	0.909
		24-months						*24-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
12-months	MLT	0.949	0.000	0.051	0.000	*12-months*	MLT	0.989	0.000	0.011	0.000
	AMN	0.000	0.938	0.031	0.031		AMN	0.000	0.979	0.000	0.021
	DYS/BA	0.000	0.000	0.946	0.054		DYS/BA	0.032	0.140	0.753	0.075
	AVG	0.000	0.062	0.000	0.938		AVG	0.000	0.082	0.000	0.918

Note: APOE = apolipoprotein E; MLT = multi-domain impairment; AMN = amnestic impairment; DYS/BA = dysexecutive/ below average cognition; AVG = average cognition

The effect of baseline CSF $pTau_{181}$ positivity on transition probabilities is presented in Table 27. MLT participants remained within class from baseline to 12-months regardless of $pTau_{181}$ positivity (100%). However, they had a lower probability to stay in their class (76%) and a higher likelihood of

transition to the DYS/BA class from 12- to 24-months (24%) with $pTau_{181}$ negativity, compared to a very high probability to remain within the MLT class from 12- to 24-months (99%) if they were $pTau_{181}$ positive. Individuals in the AMN class were highly likely to remain within class across the three time points (93–98%) regardless of $pTau_{181}$ positivity. DYS/BA class participants were most likely to remain within class from baseline to 12-months irrespective of $pTau_{181}$ positivity (94%), though these individuals were more likely to transition to the AMN (11%) and AVG (9%) classes from 12- to 24-months with $pTau_{181}$ negativity compared to those that were $pTau_{181}$ positive (AMN: 7%, AVG: 4%). $pTau_{181}$ negative individuals in the AVG class were very likely to remain within class across the three time points (90–100%), though they had a modest likelihood of transition to the AMN class from baseline to 12-months (10%). Conversely, AVG participants with $pTau_{181}$ positivity had lower probabilities to stay in the AVG class (83–88%) and a greater likelihood of transition to the AMN class from baseline to 12-months and 12- to 24-months (13%).

Table 27: Latent transition probabilities between classes from baseline to 12-months and 12- to 24-months by baseline cerebrospinal fluid $pTau_{181}$ positivity.

		$pTau_{181}$ negative						**$pTau_{181}$ positive**			
		12-months						*12-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
Baseline	MLT	1.000	0.000	0.000	0.000	*Baseline*	MLT	1.000	0.000	0.000	0.000
	AMN	0.000	0.940	0.015	0.045		AMN	0.007	0.974	0.000	0.020
	DYS/BA	0.063	0.000	0.937	0.000		DYS/BA	0.000	0.000	0.945	0.055
	AVG	0.000	0.102	0.000	0.898		AVG	0.000	0.131	0.040	0.829
		24-months						*24-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
12-months	MLT	0.757	0.000	0.243	0.000	*12-months*	MLT	0.988	0.000	0.012	0.000
	AMN	0.000	0.931	0.035	0.035		AMN	0.000	0.979	0.000	0.021
	DYS/BA	0.000	0.109	0.806	0.085		DYS/BA	0.000	0.065	0.893	0.042
	AVG	0.000	0.000	0.000	1.000		AVG	0.000	0.125	0.000	0.875

Note: $pTau_{181}$ positivity based on Schindler et al. (2018) cutoffs. $pTau_{181}$ negative $\leq$ 19.2 pg/mL, $pTau_{181}$ positive >19.2 pg/mL. $pTau_{181}$ = tau phosphorylated at amino acid-181; MLT = multi-domain impairment; AMN = amnestic impairment; DYS/BA = dysexecutive/ below average cognition; AVG = average cognition.

The effect of baseline CSF total tau positivity on transition probabilities is presented in Table 28. MLT participants were more likely to remain within class from baseline to 12-months if total tau positive

(100%) versus total tau negative (90%); total tau negativity also resulted in a modest transition probability to the DYS/BA class (10%). Individuals in the MLT class had an equally high likelihood of remaining within class from 12- to 24-months irrespective of total tau positivity. AMN participants were also more likely to stay within the AMN class from baseline to 12-months if they were total tau positive (99%) versus total tau negative (92%), and total tau negativity resulted in a small probability of transition to the AVG class (8%). Individuals in the AMN class had an equally high likelihood of remaining within class from 12- to 24-months irrespective of total tau positivity (95–96%). Total tau positive individuals in the DYS/BA class were more likely to remain within their class across the three time points (93–95%) than total tau negative DYS/BA participants (82–91%), with small and varying transition probabilities to the other three classes (3–11%) irrespective of total tau positivity. Total tau negative individuals in the AVG class were very likely to remain within class across the three time points (91–96%), with a small likelihood of transition to the AMN class (4–9%). Conversely, AVG participants with total tau positivity were less likely to stay in the AVG class (76–89%), and had modest transition probabilities to the AMN class across the three time points (11–15%) as well as a small likelihood of transition to the DYS/BA class from baseline to 12-months (9%).

Table 28: Latent transition probabilities between classes from baseline to 12-months and 12- to 24-months by baseline cerebrospinal fluid total tau positivity.

		Total tau negative						**Total tau positive**			
		12-months						*12-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
Baseline	MLT	0.896	0.000	0.104	0.000	*Baseline*	MLT	1.000	0.000	0.000	0.000
	AMN	0.000	0.924	0.000	0.076		AMN	0.013	0.987	0.000	0.000
	DYS/BA	0.091	0.000	0.909	0.000		DYS/BA	0.000	0.000	0.952	0.048
	AVG	0.000	0.090	0.000	0.910		AVG	0.000	0.149	0.093	0.758
		24-months						*24-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
12-months	MLT	0.985	0.000	0.015	0.000	*12-months*	MLT	0.989	0.000	0.011	0.000
	AMN	0.000	0.948	0.024	0.027		AMN	0.000	0.962	0.013	0.025
	DYS/BA	0.000	0.112	0.820	0.068		DYS/BA	0.000	0.042	0.928	0.030
	AVG	0.000	0.040	0.000	0.960		AVG	0.000	0.109	0.000	0.891

Note: Total tau positivity based on Schindler et al. (2018) cutoffs. Total tau negative ≤242 pg/mL, Total tau positive >242 pg/mL. MLT = multi-domain impairment; AMN = amnestic impairment; DYS/BA = dysexecutive/ below average cognition; AVG = average cognition.

The effect of baseline CSF $A\beta_{1\text{-}42}$ positivity on transition probabilities is presented in Table 29. MLT participants remained within class from baseline to 12-months regardless of $A\beta_{1\text{-}42}$ positivity (100%). However, they had a lower probability to stay in their class (62%) and a higher likelihood of transition to the DYS/BA class from 12- to 24-months (38%) if $A\beta_{1\text{-}42}$ negative, compared to a very high probability to remain within the MLT class from 12- to 24-months (99%) if they were $A\beta_{1\text{-}42}$ positive. Individuals in the AMN class were more likely to remain within class from baseline to 12-months with $A\beta_{1\text{-}42}$ positivity (99%), and had a lower probability to stay in the AMN class if $A\beta_{1\text{-}42}$ negative (88%). $A\beta_{1\text{-}42}$ negative AMN participants also had equally small likelihoods of transition to either the DYS/BA or AVG classes (6%). DYS/BA class participants were most likely to remain within class from baseline to 12-months irrespective of $A\beta_{1\text{-}42}$ positivity (91–97%), though these individuals were less likely to remain in class from 12- to 24-months if they were $A\beta_{1\text{-}42}$ negative (77%) versus $A\beta_{1\text{-}42}$ positive (88%). Regardless of $A\beta_{1\text{-}42}$ positivity, individuals in the DYS/BA class had small transition probabilities to the AMN class from 12- to 24-months (7-10%), though they had a greater likelihood of transition to the AVG class with $A\beta_{1\text{-}42}$ positivity (13%) than $A\beta_{1\text{-}42}$ negativity (5%). $A\beta_{1\text{-}42}$ positivity had little effect on the likelihood of individuals in the AVG class

remaining in their class from baseline to 12-months (85–86%) or 12- to 24-months (92–95%). Similar transition probabilities for AVG participants to the DYS/BA class were also observed irrespective of $A\beta_{1\text{-}42}$ positivity from baseline to 12-months (11–14%) and 12- to 24-months (5–8%).

Table 29: Latent transition probabilities between classes from baseline to 12-months and 12- to 24-months by baseline cerebrospinal fluid $A\beta_{1\text{-}42}$ positivity.

		$A\beta_{1\text{-}42}$ negative						**$A\beta_{1\text{-}42}$ positive**			
		12-months						*12-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
Baseline	MLT	1.000	0.000	0.000	0.000	*Baseline*	MLT	1.000	0.000	0.000	0.000
	AMN	0.000	0.877	0.060	0.063		AMN	0.005	0.995	0.000	0.000
	DYS/BA	0.032	0.000	0.968	0.000		DYS/BA	0.048	0.000	0.906	0.045
	AVG	0.000	0.113	0.033	0.853		AVG	0.000	0.141	0.000	0.859
		24-months						*24-months*			
		MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
12-months	MLT	0.624	0.000	0.376	0.000	*12-months*	MLT	0.994	0.002	0.004	0.000
	AMN	0.000	0.972	0.000	0.028		AMN	0.000	0.960	0.019	0.021
	DYS/BA	0.000	0.099	0.771	0.129		DYS/BA	0.000	0.067	0.884	0.049
	AVG	0.000	0.081	0.000	0.919		AVG	0.000	0.054	0.000	0.946

Note: $A\beta_{1\text{-}42}$ positivity based on Schindler et al. (2018) cutoffs. $A\beta_{1\text{-}42}$ negative ≥1098 pg/mL, $A\beta_{1\text{-}42}$ positive <1098 pg/mL. $A\beta_{1\text{-}42}$ = amyloid-beta$_{1\text{-}42}$; MLT = multi-domain impairment; AMN = amnestic impairment; DYS/BA = dysexecutive/ below average cognition; AVG = average cognition.

The effect of baseline ATN classification on transition probabilities is presented in Table 30. With the exception of A+/T+/N-, MLT participants remained within class from baseline to 12-months irrespective of ATN classification (100%). However, from 12- to 24-months individuals in the MLT class with A-/T-/N-, A+/T-/N-, and A-/T+/N+ classifications had a lower likelihood to stay in their class (60–86%) and a modest to large probability of transition to the DYS/BA class (12–40%). MLT participants classified as A+/T+/N- had an equal likelihood to stay in the MLT class across the three time points (85%) and a modest transition probability (15%) from baseline to 12-months to the AMN class and from 12- to 24-months to the DYS/BA class. Classification in the A+/T+/N+ resulted in a 100% probability to remain within the MLT class across the three time points.

Individuals in the AMN class had a very high likelihood to remain within their class across the three time points (94–100%) for the three ATN classifications where the "A" (i.e., $A\beta_{1-42}$) was positive (i.e., A+/T-/N-, A+/T+/N-, or A+/T+/N+). AMN participants that were A-/T-/N- had a lower probability to remain in their class (83–92%) and a small likelihood of transition to the AVG class (8–9%) across the three time points. AMN participants in the A-/T+/N+ classification had a modest likelihood to transition to the DYS/BA class (13%) from baseline to 12-months, but stayed in the AMN class from 12- to 24-months (100%).

DYS/BA participants classified as A+/T+/N+ had a high probability of remaining in their class across the three time points (95%). A-/T-/N-, A+/T-/N-, and A-/T+/N+ individuals in the DYS/BA class had a high probability of remaining in their class from baseline to 12-months (93–98%), though they had a lower likelihood of staying in class from 12- to 24-months (83–84%). From 12- to 24-months, DYS/BA participants classified as A+/T-/N- and A-/T+/N+ had a modest likelihood of transition to the AVG class (15–17%) and those in the A-/T-/N- classification had a modest transition probability to the AMN class (16%). DYS/BA individuals classified as A+/T+/N- were moderately likely to stay in their class from baseline to 12-months, and had large (45%) and modest probabilities (19%) to respectively transition to the MLT and AVG classes over this time period. However, A+/T+/N- participants in the DYS/BA class remained in the DYS/BA class from 12- to 24-months.

A-/T-/N- and A+/T+/N- individuals in the AVG class had a high probability to remain in their class (81–82%) along with a modest likelihood of transition to the DYS/BA class from baseline to 12-months (17–18%). From 12- to 24-months these the AVG participants classified as A-/T-/N- and A+/T+/N- had a very high probability to stay in the AVG class (97–100%). AVG class participants in the A+/T-/N- classification were most likely to remain within their class across the three time points (96–97%). AVG individuals classified as A-/T+/N+ had a high probability to stay within class (84–90%) and a small likelihood of transition to the DYS/BA class (9–10%) across the three time points; they also demonstrated a small transition probability to the AMN class from baseline to 12-months (7%). Although the majority of participants identified as A+/T+/N+ in the AVG class remained in their class across the three time points (68 – 88%), they also demonstrated a modest likelihood of transition to the AMN class (11-19%) from

baseline to 12-months and 12- to 24-months as well as a modest transition probability to the DYS/BA class (13%) from baseline to 12-months.

Table 30: Latent transition probabilities between classes from baseline to 12-months and 12- to 24-months by baseline A/T/N classification.

			12-months						*24-months*			
A–			MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
T–	*Baseline*	MLT	1.000	0.000	0.000	0.000	*12-months*	MLT	0.667	0.000	0.333	0.000
		AMN	0.000	0.831	0.086	0.083		AMN	0.000	0.915	0.000	0.085
N–		DYS/BA	0.066	0.000	0.934	0.000		DYS/BA	0.000	0.156	0.826	0.017
		AVG	0.000	0.172	0.023	0.805		AVG	0.000	0.000	0.000	1.000
			12-months						*24-months*			
A+			MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
T–	*Baseline*	MLT	1.000	0.000	0.000	0.000	*12-months*	MLT	0.857	0.021	0.121	0.000
		AMN	0.000	0.944	0.056	0.000		AMN	0.000	0.947	0.053	0.000
N–		DYS/BA	0.072	0.000	0.928	0.000		DYS/BA	0.015	0.000	0.839	0.146
		AVG	0.000	0.000	0.031	0.969		AVG	0.000	0.037	0.000	0.963
			12-months						*24-months*			
A+			MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
T+	*Baseline*	MLT	0.852	0.000	0.148	0.000	*12-months*	MLT	0.849	0.151	0.000	0.000
		AMN	0.000	1.000	0.000	0.000		AMN	0.000	1.000	0.000	0.000
N–		DYS/BA	0.451	0.000	0.362	0.186		DYS/BA	0.000	0.000	1.000	0.000
		AVG	0.000	0.178	0.000	0.822		AVG	0.000	0.033	0.000	0.967
			12-months						*24-months*			
A–			MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
T+	*Baseline*	MLT	1.000	0.000	0.000	0.000	*12-months*	MLT	0.601	0.000	0.399	0.000
		AMN	0.000	0.866	0.134	0.000		AMN	0.000	1.000	0.000	0.000
N+		DYS/BA	0.000	0.017	0.983	0.000		DYS/BA	0.000	0.000	0.831	0.169
		AVG	0.000	0.092	0.071	0.836		AVG	0.000	0.102	0.000	0.898
			12-months						*24-months*			
A+			MLT	AMN	DYS/BA	AVG			MLT	AMN	DYS/BA	AVG
T+	*Baseline*	MLT	1.000	0.000	0.000	0.000	*12-months*	MLT	1.000	0.000	0.000	0.000
		AMN	0.016	0.984	0.000	0.000		AMN	0.000	0.969	0.000	0.031
N+		DYS/BA	0.000	0.000	0.954	0.046		DYS/BA	0.000	0.055	0.945	0.000
		AVG	0.000	0.186	0.133	0.681		AVG	0.000	0.113	0.000	0.887

Note: A/T/N classification defined according to Jack et al. (2016). 'A' reflects amyloid ($A\beta_{1\text{-}42}$), 'T' reflects tau ($pTau_{181}$), and 'N' reflects neurodegeneration (total tau). Cerebrospinal fluid $A\beta_{1\text{-}42}$, $pTau_{181}$, and total tau positivities were based on Schindler et al. (2018) cutoffs. $A\beta_{1\text{-}42}$ = Amyloid-beta$_{1\text{-}42}$; $pTau_{181}$ = tau phosphorylated at amino acid-181; MLT = multi-domain impairment; AMN = amnestic impairment; DYS/BA = dysexecutive/ below average cognition; AVG = average cognition.

Study 1, in part, is a reprint of the material as it appears in the *Journal of the International Neuropsychological Society, (2017), 23*(7), 564-576. Eppig, Joel S.; Edmonds, Emily C.; Campbell, Laura; Sanderson-Cimino, Mark; Delano-Wood, Lisa; and Bondi, Mark W. for the Alzheimer's Disease Neuroimaging Initiative, Cambridge University Press. The dissertation author was the primary investigator and author of this paper.

V. Discussion

The present dissertation research aimed to use latent mixture modeling to empirically identify and longitudinally characterize subtypes of mild cognitive impairment. To achieve this objective, the dissertation was comprised of three inter-related studies that built upon and advanced the foundation developed by the preceding study. Study 1 used LPA to investigate baseline neuropsychological heterogeneity across four cognitive domains (visuoconstructional ability, language, attention/executive function, and episodic memory) in ADNI participants diagnosed with conventional MCI. In contrast to past neuropsychological research in ADNI, tests of visuoconstructional ability were included to better capture aspects of visuospatial functioning in statistically-defined MCI subtypes and facilitate the identification of potentially unique cognitive phenotypes with visuoconstructional deficits. Exploratory outcomes of cerebrospinal fluid and genetic AD biomarkers, longitudinal outcome, and other ADNI measures were also evaluated between LPA classes. Study 2 examined the stability and consistency of MCI subtypes across 24-months using psychometrically advanced standardized scores. To build and improve upon the normative methods of study 1, an adaptation of the Heaton, Miller, Taylor & Grant (2004) standardization methodology was used to produce demographically-corrected T-scores at baseline, 12-, and 24-months based on robust normal control performance. Practice effects were also accounted for within the final T-scores as the robust normal control participants completed all neuropsychological tests at each time point. Separate LPAs of neuropsychological performance at baseline, 12-, and 24-months were conducted on ADNI participants diagnosed with baseline MCI to establish serial measurement models. Mean performance within analogous classes and reliable change classifications between classes elucidated changes in neuropsychological performance over time. Study 3 investigated changes in class membership of the neuropsychological MCI subtypes over 24-months. Measurement invariance was evaluated within the LPA classes from study 2 and LTA was then used to determine transition probabilities from baseline to 12-months, and 12-months to 24-months. AD-risk factor covariates (i.e., AD cerebrospinal fluid and genetic biomarkers, functional ability) were added to the model to examine their effect on the likelihood of transition between classes over time.

In study 1, the optimal LPA solution contained three classes: a mixed MCI, an amnestic MCI, and LPA-derived normal class. Contrary to our expectations, a unique MCI subtype characterized by predominant visuoconstructional deficits did not emerge in the 3-class LPA. Several reasons might explain the absence, including the neuropsychological measures chosen, psychometric properties of scoring systems, selected latent model, and MCI diagnostic criteria used by ADNI.

Visuospatial assessments available in ADNI were unfortunately limited to visuoconstructional tasks, which are multi-factorial and require integration of visuoperceptual, organizational, and motor skills (Ahmed et al., 2016). Thus, low scores on clock drawing and MMSE pentagons may reflect a combination of visuospatial and executive functioning difficulties rather than "pure" visuospatial impairment. Additionally, the psychometric properties of these visuoconstructional measures were non-normally distributed, did not benefit from transformations, and poorly discriminated between normal and mildly impaired individuals (Eppig et al., 2017), likely contributing to these results.

Another possible contribution is the initial ADNI diagnosis of MCI. ADNI inclusion criteria are heavily weighted towards verbal episodic memory to target preclinical AD, while previous research has demonstrated early, differential visuospatial/constructional impairment most frequently in individuals with non-amnestic MCI (Clark et al., 2013; Ferman et al., 2013; Molano et al, 2010). Thus, one might argue that ADNI's reliance on a single memory score to determine MCI potentially biases the prevalence of non-amnestic deficits in ADNI. However, visuoconstructional impairment is not captured by verbal memory assessment and, along with other non-amnestic domains, remains uncharacterized with ADNI's diagnostic criteria. In fact, past work (Bondi et al., 2014; Edmonds et al., 2015) has demonstrated considerable heterogeneity in ADNI neuropsychological profiles despite the vast majority of individuals receiving a conventional "amnestic MCI" diagnosis. Furthermore, recent research also indicates that the "pure" AD pathology targeted by ADNI is less common than multiple underlying neuropathologies (Schneider et al., 2009; Wilson et al., 2013; Zlokovic, 2011), providing further support for using comprehensive neuropsychological assessment to classify MCI across multiple cognitive domains.

Unsurprisingly, results of the study 1 are similar to the cluster subgroups found by Edmonds et al. (2015), who reported analogous amnestic MCI (34.9%), dysexecutive MCI (12.5%), and cluster-derived

normal (34.2%) subtypes. Although amnestic MCI class in study 1 was much larger (56.5%), the LPA-derived normal class was comparable in size (30.4%). The mixed MCI class (13.2%) appears to correspond to the dysexecutive MCI group in Edmonds et al. (2015), with analogous size and performance. However, Edmonds et al. (2015) also found a fourth dysnomic/amnestic MCI subtype. There are two possibilities explaining its absence in study 1: 1) The statistical algorithms underlying LPA, which converged on a different solution, and 2) inclusion of visuoconstructional assessment, which revealed more robust impairment in a subset of dysnomic individuals. These participants may have been reclassified as mixed MCI in the LPA and the remaining dysnomic subjects, lacking adequate differentiation in their scores, were folded into the amnestic class. Additionally, the 3-class solution was very consistent with another ADNI cluster analysis by Bondi et al. (2014) using conventional Petersen/Winblad MCI criteria. They also found three MCI subgroups (amnestic: 56.4%, dysexecutive/mixed: 12.3%, and "false positive" normal: 31.3%) of almost identical size and cognitive profile, along with similar genetic/CSF biomarker associations and longitudinal outcomes.

Overall, study 1 is very consistent with previous findings and further underscores the problem using single test scores, cognitive screening measures, and subjective rating scales in MCI diagnosis. ADNI's MCI criteria led to "false-positive" diagnoses in approximately a third of the sample; this class performed within-normal limits on all neuropsychological measures, had a lower proportion of AD-positive CSF and genetic biomarkers, and better longitudinal outcomes compared to the other MCI classes, similar to past results (Bondi et al., 2014; Edmonds et al., 2015). Although the amnestic and mixed MCI classes in study 1 demonstrated unique neuropsychological profiles, the groups only differed in total MMSE score and rate of conversion to AD among all CSF/genetic biomarker, ADNI diagnostic, and longitudinal outcomes. These results raise the possibility that the amnestic and mixed MCI classes may represent stages of disease progression. A recent analysis of cortical atrophy patterns among Edmond et al.'s (2015) cluster-defined MCI subtypes supports such speculation, as the authors demonstrated distinct but overlapping profiles of cortical thinning consistent with their neuropsychological performance (Edmonds et al., 2016).

Despite the majority of similarities, a few notable differences were present in this study compared to past work (Bondi et al., 2014; Edmonds et al., 2015). Most importantly, the LPA-derived normal class yielded a rate of dementia progression (5.8%) that is almost half of Edmonds et al.'s (2015) finding (10.7%) and Bondi et al.'s (2014) results (9.3%). This significant result suggests LPA methods in study 1 further improved classification accuracy and are preferential to cluster analysis. Another unique finding in study 1 was the disproportionate representation of the LPA-derived normal class by ADNI phase; fewer such individuals were enrolled during ADNI-1 than other classes, though significantly more were recruited in ADNI-GO. This shift likely reflects ADNI-GO's efforts to focus on "early" MCI (Aisen et al., 2010). However, without the incorporation of comprehensive neuropsychological assessment to inform diagnosis, ADNI may have unintentionally recruited cognitively normal individuals erroneously identified as "early" MCI. Such misclassification has considerable implications for MCI research, where inaccurate diagnosis will increase the likelihood of Type-II errors, attenuate effects sizes, and reduce the efficacy of pharmacologic interventions (e.g., see Edmonds et al., 2018).

In study 2, application of the Heaton, Miller, Taylor, & Grant (2004) standardization methodology to robust normal control performance at baseline, 12-, and 24-months resulted in demographically-corrected T-scores with satisfactory psychometric properties and embedded practice effects. As predicted, all T-scores had a mean of 50 and standard deviation of 10 in the robust normal control sample, and the percentage of scores in the "impaired" range (i.e., T<40) amongst robust normal control participants did not statistically differ from the percentage of individuals expected to fall within that range (i.e., 15.9%) on the normal curve. As expected in the robust normal control sample, there were no residual associations between age, education, and sex and the final T-scores, or an age by education interaction. Most of the frequency distributions in the robust normal control sample for each T-score did not differ from a normal curve. However, contrary to expectations the distributions of BNT and AVLT Recognition at all three time points were significantly different than the normal curve. A likely explanation would be that despite the total range of the measures, the majority of healthy individuals actually perform within a narrow band on these tests, thus reflecting a different distribution. This notion has been

supported in the normative literature, where other measures with limited ranges do not tend to adhere to a Gaussian distribution (Heaton, Miller, Taylor, & Grant, 2004).

Contrary to study 1 results, in study 2 the optimal LPA solutions at baseline, 12-months, and 24-months produced four analogous classes: a multi-domain impairment, an amnestic impairment, a dysexecutive/below average cognition, and an average cognition class. Several reasons for the differences in the number of classes between study 1 and 2 are possible, including the exclusion of visuoconstructional measures and improved psychometric properties of standardized scores, allowing for more robust determination of classes. Additionally, comparative fit indices, likelihood ratio tests, model characteristics, and parsimony did not all converge on a clear solution for the number of LPA classes at baseline, 12- and 24-months. Rather, a combination of these objective statistical indices along with subjective factors such as meaningful class interpretation and consistency with past research were used to select the optimal model at all three time points. Thus, another reason for the difference in the study 1 LPA versus the study 2 LPAs is the possibility that the incorrect class solution was selected, although objective indicators of fit (e.g., VLMR-LRT) suggest that class over-extraction is unlikely.

Another reason that does not support data over-extraction is that the 4-class LPA solutions in study 2 are very consistent with the 4-classes found in past cluster analytic studies of conventional MCI (Edmonds et al., 2015; Bondi et al., 2014; Peraita et al., 2015). Similar to the amnestic, dysexecutive/mixed, and cognitively normal classes of Edmonds et al. (2015), study 2 found analogous amnestic impairment, multi-domain impairment, and average cognition classes in ADNI, albeit of different sizes. One notable difference between study 2 and their work is the size of the “false-positive” cognitively normal class, as it is 2–3x larger (depending on the time point chosen from study 2) in Edmonds et al. (2015). The class size of the average cognition class in study 2 varied greatly across time points, and in conjunction with better performance and the larger size of the dysexecutive/below average cognition class at 12- and 24-months, may indicate that the LPAs in study 2 are underestimating the size of the “false-positive” subtype as some participants were classified into a different group at later time points. Another difference from Edmonds et al. (2015) is the interpretation of the 4th class; the authors reported a dysnomic/amnestic subtype, while study 2 found a dysexecutive/below average cognition subtype. The

exact reason for this difference is unclear, although it may be due to the statistical method as well as the general variability in performance of participants in the dysexecutive/below average class over time.

The neuropsychological profiles of LPA classes at each of the three time points in study 2 also appears to be relatively similar to those found in serial LPAs across 24-months by Peraita et al. (2015). These authors consistently produced 4-classes at three time points, each containing a healthy non-impairment class, amnesic class, non-amnesic class, and multi-domain class which are analogous to the average cognition, amnestic impairment, dysexecutive/below average cognition, and multi-domain impairment classes in study 2. Similar to the findings in study 2, Peraita et al. (2015) also demonstrated that the cognitively normal class decreased in size with time, although the proportion of individuals in this class was 5–7x in Peraita et al. (2015). One possible explanation for this discrepancy between our studies is the use of continuous versus dichotomous indicators. Further compounding this difference was Peraita et al.'s (2015) use of a very liberal classification of "impairment" (<40th percentile), which is likely to be over inclusive of healthy normal individuals. Study 2 included continuous neuropsychological indicators of latent classes, thus avoiding unnecessary assumptions or dichotomizations of measures.

Although the number of classes remained constant across the 24-months as predicted, the sizes of the classes, particularly the dysexecutive/below average cognition and average cognition subtypes, had more fluctuation than anticipated. Moreover, the dysexecutive/below average cognition class had improvement in neuropsychological performance over the 24-months. The reasons underlying the variations in size and to an extent, neuropsychological interpretation, of the dysexecutive/below average cognition class is not immediately evident. However, one possible explanation may relate to the relationship between missing data at follow-up time points and participants in this class. Although an analysis of missing follow-up data did not statistically differ across the baseline or 12-month LPA classes, practical examination shows that the dysexecutive/below average cognition class did have the highest percentage of missing data, particularly for 24-months follow-up in the baseline class. Additionally, MCI participants in the overall sample that were missing data at 24-months tended to perform worse on the attention/executive function measures (TMT Part A & B) than those with available follow-up. The combination of these two factors indicates the possibility that there was an association between

individuals with missing follow-up data and having worse performance on attention/executive function measures. Thus, their absence in the 12- and 24-month LPAs may have disproportionately affected the dysexecutive/below average cognition class and caused greater fluctuations in its size and improvement on these neuropsychological measures over time. As expected, within-class comparisons suggested the multi-domain impairment class performed worse on virtually all measures across the three time points. The amnestic impairment class had decreased neuropsychological performance on several measures over time, while performance within the average cognition class remained stable as predicted.

Examination of outcomes between classes for each LPA suggested changing demographic differences across time points, which is likely a function of the changes in participant composition and size of classes at baseline, 12-, and 24-months. At the baseline and 24-month LPAs, the average cognition class had a tendency to be younger than some of the other classes, and the multi-domain impairment class generally had a higher percentage of women than the amnestic impairment and dysexecutive/below average cognition classes. Furthermore, the multi-domain impairment class had the highest proportion of participants with a concurrent dementia diagnosis from ADNI compared to the other classes in the 12- and 24-month LPAs; up to 73% received a dementia diagnosis at 24-months. The average cognition class had a higher proportion of individuals that were identified by ADNI as normal at their concurrent 12- and 24-month LPAs. These findings provide further support to evidence from past research suggesting that the average cognition class reflects cognitively normal individuals due to the susceptibility of conventional MCI to “false-positive” diagnoses. Additionally, the combination of similar neuropsychological interpretation of the multi-domain impairment class from baseline to 24-months, in conjunction with the high percentages of individuals within this class that received a dementia diagnosis, suggests that these participants may have also been misclassified with MCI at baseline and actually reflected a false-positive dementia diagnosis.

This notion is further supported based on the results of differences in reliable change classification between LPA classes at 12- and 24-months. The multi-domain impairment class had the highest percentage of individuals with a significant decrease across almost all neuropsychological measures as compared to the other classes. However, AVLT recall appeared to be disproportionately

stable in reliable classification, which is consistent with the aforementioned lack of within-class mean decline on this measure in the multi-domain impairment class. The likely explanation for this stability in AVLT recall is a floor effect, such that the majority of individuals in the multi-domain impairment class already obtained the lowest possible score. The amnestic impairment class also had a large proportion of participants with a significant decrease in reliable classification compared to the dysexecutive/below average cognition and average cognition classes, though this finding was circumscribed to AVLT recognition, one of the memory measures. The proportion of individuals in the average cognition class with a significant decrease in reliable classification was consistent with that expected in a normative sample (i.e., 5%), and they also demonstrated higher percentages of a significant increase compared to the dysexecutive/below average group in the 12-month LPA. Taken together, these results continue to provide substantial evidence that the conventional MCI criteria are prone to misdiagnosis of healthy normal individuals, as well as individuals that have already progressed to the early stages of dementia.

Study 3 found support for partial measurement invariance over time based on the LPA models from study 2. In line with expectations, good model fit was found when the mean of each neuropsychological measure (except for AVLT recall) at baseline, 12-, and 24-months was held equal across time in the average cognition class. Contrary to hypotheses, in the multi-domain impairment class model fit did not support equivalent means across time on almost all of the neuropsychological measures. Similarly, model fit indicated that the dysexecutive/below average cognition and amnestic impairment classes also had equivalent means across time for very few of the neuropsychological measures.

The latent transition analysis in study 3 generated a higher degree of consistency within analogous classes across the three time points compared to study 2. Analogous classes were closer in size and demonstrated minimal fluctuation at baseline, 12-, and 24-months. Additionally, there was greater equivalency in neuropsychological interpretation of classes over time, as scores largely fell within mainly the same Heaton, Miller, Taylor, & Grant (2004) ranges. This is due, in part, to the constraint of select neuropsychological means across time within classes. However, another explanation may be related to the imputation of missing data used by MPlus (Muthén & Muthén, 2012). Bayesian imputation of follow-up scores based on the neuropsychological performance from prior time points allowed MPlus to

use all MCI participant data (n=825) at 12- and 24-months, and may have helped with the stability of solutions by accounting for those individuals with worse attention/executive performance that would otherwise be missing. However, the exploratory LPAs that were run separately in study 2 did not benefit from this form of imputation, as no relationships were established with prior time points in those models.

Overall, transition probabilities established in study 3 indicated that a significant majority of participants (>86%) remained within their class from baseline to 12-months and 12- to 24-months. As expected, individuals in the multi-domain impairment class had the highest probability to stay in their class over time (>97%). Contrary to predictions, participants in the amnestic impairment class were also very likely to remain in their class (93%) and demonstrated little to no transition to the multi-impairment class. Probability staying in class for the dysexecutive/below average class was higher from baseline to 12-months and was relatively lower from 12- to 24-months (with a slight increase in likelihood to the amnestic impairment or average cognition classes). Additionally, the dysexecutive/below average class also had the largest likelihood (although relatively small in absolute terms) of transition to the multi-domain impairment class compared to the other classes. Despite expectations, the average class had the lowest likelihood to remain in class from baseline to 12-months (though still high in absolute terms), with small probabilities of transition to the amnestic impairment and dysexecutive/below average cognition classes. However, participants in the average cognition class were more likely to remain within class from 12- to 24-months, with a small chance of transition to the amnestic impairment class. Inclusion of a second-order effect had minimal impact on the transition probabilities, an indication of little direct lasting effect of baseline class membership on 24-month class membership.

Regarding the effect of covariates on transition probabilities, poor functional ability and AD-CSF biomarkers appeared to have the largest influence on the likelihood of transition. Presence or absence of the APOE ε4 allele made little impact in the multi-domain impairment, amnestic impairment, and average cognition classes. However, the dysexecutive/below average cognition class had a slightly higher likelihood of transition to the amnestic class from 12- to 24-months in participants that were APOE ε4-positive. As anticipated, AD-CSF biomarker positivity and poor functional ability had little impact on the transition probabilities of the multi-domain class participants, who continued to remain most likely to stay

in their class over time. However, AD-CSF biomarker negativity as well as intact functional ability did increase the likelihood of transition to the dysexecutive/below average cognition class for several time points. A similar pattern was observed for A/T/N classification; in the presence of A+/T+/N+ participants in the multi-domain impairment class remained within their class. However, with two or fewer positivities in the A/T/N scheme, there were greater transition probabilities from the multi-domain impairment to the amnestic impairment and dysexecutive/below average cognition classes.

As predicted, poor functional ability in the amnestic impairment class modestly increased the likelihood of transition to the multi-domain impairment class from baseline to 12-months. However, contrary to expectations AD-positivity on CSF biomarkers had little impact on the transition probabilities of the amnestic to multi-domain impairment class. One possibility for this finding is that these LTA classes represent stages of illness progression. Thus, the lack of any intermediate stage of impairment to transition to in-between the amnestic and multi-domain subtypes may facilitate participants in the amnestic impairment class staying within their class and slowly declining in performance over time rather than suddenly transitioning to a profile of mild to moderate impairment across all cognitive domains.

Poor functional ability increased the likelihood of transition for the dysexecutive/below average cognition class to the multi-domain impairment class from baseline to 12-months and to the amnestic impairment class from 12- to 24-months. The dysexecutive/below average cognition class also had increased probability of transition to the average cognition class with CSF $A\beta_{1\text{-}42}$ negativity. Despite small variable transition probabilities to the other classes at various time points, there were few other strong patterns of transition in the dysexecutive/below average cognition class. Past research conceptualization of “non-amnestic” MCI subtypes has indicated that this group may reflect multiple or heterogeneous etiologies (Petersen et al., 2004; Winblad et al., 2004), which may contribute to the lack of a clear pattern in transition probabilities among these participants.

There was minimal impact of functional ability on the transition probabilities of the average cognition class. A similar pattern emerged with regard to the influence of $A\beta_{1\text{-}42}$. However, positivity on CSF total tau and $pTau_{181}$ increased the likelihood of transition for participants in the average cognition class to the dysexecutive/below average cognition class. The minimal role $A\beta_{1\text{-}42}$ versus the prominent

impact of both tau markers on likelihood of transition in this group may be due to several reasons. One explanation is that the ordered, sequential progression of neuropathology and neuronal loss in AD that has been promulgated by the amyloid cascade hypothesis (Jack et al., 2016) does not accurately capture potential differences within AD of the driving neuropathology. Another possibility is that the increased transition probabilities within the context of tau for individuals in the average cognition class may reflect progression of a non-AD etiology, such as primary age related tauopathy (Crary et al., 2014), that is underlying changes in cognitive profiles, although Braak & Del Tredici (2011) counter that medial temporal tauopathy typically appears well prior to the emergence of amyloidosis in the developmental continuum leading to AD. Regardless, support for agnosticism in the initial appearance of amyloid or tau changes may be found in the effect of A/T/N classification on transition probabilities; irrespective of positivity specific to $A\beta_{1\text{-}42}$ or a tau, positivity on two or more biomarkers demonstrated modest increases in likelihood of transition to the amnestic impairment and dysexecutive/below average cognition classes across the three time points. This latter 'tally' system of biomarker risks concurs with Edmonds et al.'s (2015) finding of increasing rates of progression with increasing numbers of biomarker and 'subtle cognitive decline' positivities. Taken together, AD-risk factor covariates produced unique patterns of influence on transition probabilities among the four MCI classes, with modest to large increases in the likelihood of moving to another class across 24-months. The findings presented in this dissertation may help inform efforts to codify biomarker and cognitive risks in definitional schemes of Alzheimer's disease and related disorders (e.g., see Jack et al., 2018).

Strengths and Limitations

Strengths of the current dissertation research include its large sample size, neuropsychological representation of several major cognitive domains, use of follow-up data across 24-months to examine longitudinal performance, inclusion of CSF AD-biomarkers and APOE $\varepsilon 4$ genotyping, adaptation of psychometrically advanced and well-established standardization methods to generate demographically-corrected T-scores with embedded practice effects across three time points, use of a robust normal control group with complete neuropsychological data over 24-months to generate norms, and creation of reliable change classifications between three time points. Additionally, mixture models provide a

statistically sophisticated method to classify individuals, with increasing use in the MCI classification literature (Hanfelt et al., 2011; Köhler et al., 2013; McGuiness et al., 2015). This dissertation research is also among the first to employ latent transition analysis to examine the relationships of empirical neuropsychological subtypes in MCI over time. Limitations of the current dissertation include the paucity of neuropsychological measures administered across all ADNI phases, the lack of a diverse set of visuospatial measures in ADNI, the oversampling of normal control participants with a high levels of education and the undersampling of these same participants with low levels of education (<11 years), the limited range of performance on a few neuropsychological measures in a normative sample, missing data at follow-up time points and possible associations with level of performance on some neuropsychological measures, the lack of clear convergence among objective LPA fit indices on a best-fitting model across all three time points, and the potential for small cell size in the LTA with the use of certain covariates. Additionally, LPA and LTA are unable to answer questions regarding the rate of change over time with regard to longitudinal cognitive decline in MCI subtypes.

Conclusions

Since the construct of mild cognitive impairment was introduced in the 1980's, the definition of MCI has evolved over time from its early conceptual framework put forth by Petersen and colleagues (Petersen et al., 1999; Petersen, 2004; Winblad et al., 2004) to the current diagnostic criteria (Albert et al., 2011) that better reflect the cognitive and neuropathological heterogeneity that has been increasingly recognized within MCI (Bondi et al., 2014; Clark et al., 2010; Delano-Wood et al., 2009; Edmonds et al., 2015, 2016, 2018; Eppig et al.; 2017, Libon et al., 2010; Hanfelt et al., 2011; McGuiness et al., 2015). Despite these improvements, issues remain with the ways in which conventional diagnostic criteria of MCI are operationalized in research and clinical practice. The current dissertation research adds to a growing body of literature that suggests a sizeable minority of individuals may represent false-positive diagnostic errors, the majority of whom continue to remain cognitively normal across 2 years. Additionally, this research indicates that at repeated clinical assessments, baseline MCI participants are most likely to re-produce a similar cognitive profile over time that is reflective of their initial neuropsychological MCI subtype. However, AD-CSF biomarkers and functional ability have variable effects on the likelihood of

transition over time depending on the MCI subtype. Thoughtful incorporation of these AD risk factors into baseline clinical evaluations in the future may help clinicians better understand the probability of observing a change in an individual's pattern of performance at future evaluations. Taken together, results of this dissertation research challenge the empirical validity of the conventional MCI classification system and advocate for the comprehensive neuropsychological assessment and actuarial approaches in the clinical and research diagnosis of MCI to improve classification, associations with AD biomarkers, and longitudinal outcomes (Bondi et al., 2014). Future research should examine the impact of cognitively normal "false-positives" on the attenuation of effect sizes in research studies and clinical trials, investigate the impact of neuroimaging markers of cortical thickness and AD-positive PET scans on transition probabilities, and model emergent MCI subtypes identified via actuarial neuropsychological methods that may improve diagnostic accuracy with latent mixture models.

Study 1, in part, is a reprint of the material as it appears in the *Journal of the International Neuropsychological Society, (2017), 23*(7), 564-576. Eppig, Joel S.; Edmonds, Emily C.; Campbell, Laura; Sanderson-Cimino, Mark; Delano-Wood, Lisa; and Bondi, Mark W. for the Alzheimer's Disease Neuroimaging Initiative, Cambridge University Press. The dissertation author was the primary investigator and author of this paper.

References

Ahmed, S., Brennan, L., Eppig, J., Price, C. C., Lamar, M., Delano-Wood, L., ... & Jak, A. (2016). Visuoconstructional impairment in subtypes of mild cognitive impairment. *Applied Neuropsychology: Adult*, *23*(1), 43-52. doi:10.1080/23279095.2014.1003067

Aisen, P. S., Petersen, R. C., Donohue, M. C., Gamst, A., Raman, R., Thomas, R. G., ... & Jack, C. R. (2010). Clinical core of the Alzheimer's Disease Neuroimaging Initiative: Progress and plans. *Alzheimer's & Dementia*, *6*(3), 239-246. doi:10.1016/j.jalz.2010.03.006

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., ... & Snyder, P. J. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*, *7*(3), 270-279. doi: https://doi.org/10.1016/j.jalz.2011.03.008

Alzheimer's Association. (2017). 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *13*(4), 325-373.

Alzheimer's Disease Neuroimaging Initiative. (2008). *ADNI 2 Procedures Manual.* Retrieved from http://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf

Asparouhouv, T., & Muthén, B. (2007). Wald test of mean equality for potential latent class predictors in mixture modeling. *Online technical appendix.* Los Angeles, CA: Muthén & Muthén. Available for download at https://www.statmodel.com/download/MeanTest1.pdf.

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 329-341. doi:10.1080/10705511.2014.915181

Asparouhov, T., & Muthén, B. (2015). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus Web Notes*, *21*(2), 1-22. Available for download at https://www.statmodel.com/examples/webnotes/webnote21.pdf

Attix, D. K., Story, T. J., Chelune, G. J., Ball, J. D., Stutts, M. L., Hart, R. P., & Barth, J. T. (2009). The prediction of change: Normative neuropsychological trajectories. *The Clinical Neuropsychologist*, *23*(1), 21-38. doi:10.1080/13854040801945078

Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 20-31. doi:10.1080/10705511.2014.955104

Berlin, K. S., Parra, G. R., & Williams, N. A. (2014). An introduction to latent variable mixture modeling (part 2): Longitudinal latent class growth analysis and growth mixture models. *Journal of Pediatric Psychology*, *39*(2), 188-203. doi:10.1093/jpepsy/jst085

Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An introduction to latent variable mixture modeling (part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology*, *39*(2), 174-187. doi:10.1093/jpepsy/jst084

Bittner, T., Zetterberg, H., Teunissen, C. E., Ostlund Jr, R. E., Militello, M., Andreasson, U., Hubeek, I., Gibson, D., Chu, D. C., Eichenlaub, U., Heiss, P., Kobold, U., Leinenbach, A., Madin, K.,

Manuilova, E., Rabe, C., & Blennow, K. (2016). Technical performance of a novel, fully automated electrochemiluminescence immunoassay for the quantitation of β-amyloid (1–42) in human cerebrospinal fluid. *Alzheimer's & Dementia, 12*(5), 517-526. doi: 10.1016/j.jalz.2015.09.009

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*(1), 3-27. doi:10.1093/pan/mph001

Bondi, M. W., Edmonds, E. C., Jak, A. J., Clark, L. R., Delano-Wood, L., McDonald, C. R., ... & Salmon, D. P. (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *Journal of Alzheimer's Disease*, *42*(1), 275-289. doi:10.3233/JAD-140276.

Braak, H., & Del Tredici, K. (2011). Alzheimer's pathogenesis: is there neuron-to-neuron propagation? *Acta neuropathologica*, *121*(5), 589-595. doi:10.1007/s00401-010-0789-4

Bray, B. C., Lanza, S. T., & Tan, X. (2015). Eliminating bias in classify-analyze approaches for latent class analysis. *Structural equation modeling: a multidisciplinary journal*, *22*(1), 1-11. doi:10.1080/10705511.2014.935265

Bryant, F. B. & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(3), 372-398. doi:10.1080/10705511.2012.687671

Clark, L. R., Delano-Wood, L., Libon, D. J., McDonald, C. R., Nation, D. A., Bangen, K. J., ... & Bondi, M. W. (2013). Are empirically-derived subtypes of mild cognitive impairment consistent with conventional subtypes? *Journal of the International Neuropsychological Society*, *19*(06), 635-645. doi:10.1017/S1355617713000313

Clark, S. L., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. *Submitted for publication*. Available for download at https://www.statmodel.com/download/relatinglca.pdf

Clark, S. L., Muthén, B., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural equation modeling: A multidisciplinary journal*, *20*(4), 681-703. doi:10.1080/10705511.2013.824786

Collins, L. M., & Lanza, S. T. (2013). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Hoboken, NJ: John Wiley & Sons.

Crary, J. F., Trojanowski, J. Q., Schneider, J. A., Abisambra, J. F., Abner, E. L., Alafuzoff, I., ... & Cairns, N. J. (2014). Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta neuropathologica*, *128*(6), 755-766. doi: 10.1007/s00401-014-1349-0

Crutch, S. J., Lehmann, M., Schott, J. M., Rabinovici, G. D., Rossor, M. N., & Fox, N. C. (2012). Posterior cortical atrophy. *The Lancet Neurology*, *11*(2), 170-178. doi:10.1016/S1474-4422(11)70289-7

Crutch, S. J., Schott, J. M., Rabinovici, G. D., Boeve, B. F., Cappa, S. F., Dickerson, B. C., ... & Mendez, M. F. (2013). Shining a light on posterior cortical atrophy. *Alzheimer's & Dementia*, *9*(4), 463-465. doi:10.1016/j.jalz.2012.11.004

Cummings, J. L., Morstorf, T., & Zhong, K. (2014). Alzheimer's disease drug-development pipeline: few candidates, frequent failures. Alzheimer's research & therapy, 6(4), 37. doi:10.1186/alzrt269

Cysique, L. A., Franklin Jr, D., Abramson, I., Ellis, R. J., Letendre, S., Collier, A., ... & Simpson, D. (2011). Normative data and validation of a regression based summary score for assessing meaningful neuropsychological change. *Journal of clinical and experimental neuropsychology*, *33*(5), 505-522. doi:10.1080/13803395.2010.535504

Dajani, D. R., Llabre, M. M., Nebel, M. B., Mostofsky, S. H., & Uddin, L. Q. (2016). Heterogeneity of executive functions among comorbid neurodevelopmental disorders. *Scientific reports*, *6*, 365-366. doi:10.1038/srep36566

Delano-Wood, L., Bondi, M. W., Sacco, J., Abeles, N. O. R. M., Jak, A. J., Libon, D. J., & Bozoki, A. (2009). Heterogeneity in mild cognitive impairment: Differences in neuropsychological profile and associated white matter lesion pathology. *Journal of the International Neuropsychological Society*, *15*(6), 906-914. doi:10.1017/S1355617709990257

De Santi, S., Pirraglia, E., Barr, W., Babb, J., Williams, S., Rogers, K., ... & Ferris, S. (2008). Robust and conventional neuropsychological norms: diagnosis and prediction of age-related cognitive decline. *Neuropsychology*, *22*(4), 469-484. doi:10.1037/0894-4105.22.4.469

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

Edmonds, E. C., Ard, M. C., Edland, S. D., Galasko, D. R., Salmon, D. P., & Bondi, M. W. (2018). Unmasking the benefits of donepezil via psychometrically precise identification of mild cognitive impairment: A secondary analysis of the ADCS vitamin E and donepezil in MCI study. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *4*, 11-18. doi:10.1016/j.trci.2017.11.001

Edmonds, E. C., Delano-Wood, L., Clark, L. R., Jak, A. J., Nation, D. A., McDonald, C. R., ... & Alzheimer's Disease Neuroimaging Initiative. (2015). Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. *Alzheimer's & Dementia*, *11*(4), 415-424. doi:10.1016/j.jalz.2014.03.005

Edmonds, E. C., Delano-Wood, L., Galasko, D. R., Salmon, D. P., & Bondi, M. W. (2015). Subtle cognitive decline and biomarker staging in preclinical Alzheimer's disease. *Journal of Alzheimer's disease*, *47*(1), 231-242. doi: 10.3233/JAD-150128

Edmonds, E. C., Eppig, J., Bondi, M. W., Leyden, K. M., Goodwin, B., Delano-Wood, L., ... & Alzheimer's Disease Neuroimaging Initiative. (2016). Heterogeneous cortical atrophy patterns in MCI not captured by conventional diagnostic criteria. *Neurology*, *87*(20), 2108-2116. doi:10.1212/WNL.0000000000003326

Eppig, J. S., Edmonds, E. C., Campbell, L., Sanderson-Cimino, M., Delano-Wood, L., Bondi, M. W., & Alzheimer's Disease Neuroimaging Initiative. (2017). Statistically derived subtypes and associations with cerebrospinal fluid and genetic biomarkers in mild cognitive impairment: A latent profile analysis. *Journal of the International Neuropsychological Society*, *23*(7), 564-576. doi:10.1017/S135561771700039X

Ferman, T. J., Smith, G. E., Boeve, B. F., Graff-Radford, N. R., Lucas, J. A., Knopman, D. S., ... & Dickson, D. W. (2006). Neuropsychological differentiation of dementia with Lewy bodies from

normal aging and Alzheimer's disease. *The Clinical Neuropsychologist*, *20*(4), 623-636. doi:10.1080/13854040500376831

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

Finch, H. (2015). A Comparison of Statistics for Assessing Model Invariance in Latent Class Analysis. *Open Journal of Statistics*, *5*(3), 191-210. doi:10.4236/ojs.2015.53022

Freedman, L., & Dexter, L. E. (1991). Visuospatial ability in cortical dementia. *Journal of clinical and experimental neuropsychology*, *13*(5), 677-690. doi:10.1080/01688639108401082

Geldmacher, D. S. (2003). Visuospatial dysfunction in the neurodegenerative diseases. *Frontiers in bioscience: A journal and virtual library*, *8*, e428-436. doi:10.2741/1143

Gladsjo, J. A., Schuman, C. C., Evans, J. D., Peavy, G. M., Miller, S. W., & Heaton, R. K. (1999). Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment*, *6*(2), 147-178. doi:10.1177/107319119900600204

Goodglass, H. & Kaplan, E. (1983). *The assessment of aphasia and related disorders.* Philadelphia, PA: Lea & Febiger.

Graham, J. W., Collins, L. M., Wugalter, S. E., Chung, N. K., & Hansen, W. B. (1991). Modeling transitions in latent stage-sequential processes: a substance use prevention example. *Journal of consulting and clinical psychology*, *59*(1), 48-57. doi:10.1037/0022-006X.59.1.48

Grossi, D., & Trojano, L. (2001). Constructional and visuospatial disorders. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 4, pp. 99 – 120). Amsterdam, Netherlands: Elsevier Science, B.V.

Gu, J. (2016). *Monitoring the progression of Alzheimer's disease with latent transition models.* (Master's Thesis, Kansas State University). Retrieved from http://krex.k-state.edu/dspace/bitstream/handle/2097/32919/JienaGu2016.pdf?sequence=1

Hamilton, J. M., Salmon, D. P., Galasko, D., Raman, R., Emond, J., Hansen, L. A., ... & Thal, L. J. (2008). Visuospatial deficits predict rate of cognitive decline in autopsy-verified dementia with Lewy bodies. *Neuropsychology*,*22*(6), 729-737. doi:10.1037/a0012949

Hanfelt, J. J., Wuu, J., Sollinger, A. B., Greenaway, M. C., Lah, J. J., Levey, A. I., & Goldstein, F. C. (2011). An exploration of subgroups of mild cognitive impairment based on cognitive, neuropsychiatric and functional features: Analysis of data from the National Alzheimer's Coordinating Center. *The American Journal of Geriatric Psychiatry*, *19*(11), 940-950. doi:10.1097/JGP.0b013e31820ee9d2

Harrington, K. D., Lim, Y. Y., Ames, D., Hassenstab, J., Rainey-Smith, S., Robertson, J., ... & Maruff, P. (2017). Using robust normative data to investigate the neuropsychology of cognitive aging. *Archives of Clinical Neuropsychology*, *32*(2), 142-154. doi:10.1093/arclin/acw106

Hayden, K. M., Kuchibhatla, M., Romero, H. R., Plassman, B. L., Burke, J. R., Browndyke, J. N., & Welsh-Bohmer, K. A. (2014). Pre-clinical cognitive phenotypes for Alzheimer disease: a latent profile approach. *The American Journal of Geriatric Psychiatry*, *22*(11), 1364-1374. doi:10.1016/j.jagp.2013.07.008

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead-Reitan battery: Demographic corrections, research findings, and clinical applications; with a supplement for the Wechsler Adult Intelligence Scale-Revised (WAIS-R)*. Lutz, FL: Psychological Assessment Resources.

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults.* Lutz, FL: Psychological Assessment Resources.

Heaton, R. K., Temkin, N., Dikmen, S., Avitable, N., Taylor, M. J., Marcotte, T. D., & Grant, I. (2001). Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Archives of clinical neuropsychology*,*16*(1), 75-91. doi:10.1016/S0887-6177(99)00062-1

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological methods, 11*, 36–53. doi:10.1037/1082-989X.11.1.36

Holtzer, R., Goldin, Y., Zimmerman, M., Katz, M., Buschke, H., & Lipton, R. B. (2008). Robust norms for selected neuropsychological tests in older adults. *Archives of Clinical Neuropsychology*, *23*(5), 531-541. doi:10.1016/j.acn.2008.05.004

Jack Jr, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Liu, E. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535-562. doi: 10.1016/j.jalz.2018.02.018

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., ... & Petersen, R. C. (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer's disease biomarkers. *Neurology*, *87*(5), 539-547. doi:10.1212/WNL.0000000000002923

Jak, A. J., Bondi, M. W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D. P., & Delis, D. C. (2009). Quantification of five neuropsychological approaches to defining mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, *17*(5), 368-375. doi:10.1097/JGP.0b013e31819431d5

Jefferson, A. L., Cosentino, S. A., Ball, S. K., Bogdanoff, B., Leopold, N., Kaplan, E., & Libon, D. J. (2002). Errors produced on the mini-mental state examination and neuropsychological test performance in Alzheimer's disease, ischemic vascular dementia, and Parkinson's disease. *The Journal of neuropsychiatry and clinical neurosciences*, 14(3), 311-320. doi:10.1176/jnp.14.3.311

Johnson, D. K., Morris, J. C., & Galvin, J. E. (2005). Verbal and visuospatial deficits in dementia with Lewy bodies. *Neurology*, *65*(8), 1232-1238. doi:10.1212/01.wnl.0000180964.60708.c2

Kao, A. W., Racine, C. A., Quitania, L. C., Kramer, J. H., Christine, C. W., & Miller, B. L. (2009). Cognitive and neuropsychiatric profile of the synucleinopathies: Parkinson's disease, dementia with Lewy bodies and multiple system atrophy. *Alzheimer disease and associated disorders*, *23*(4), 365-370. doi:10.1097/WAD.0b013e3181b5065d

Köhler, S., Hamel, R., Sistermans, N., Koene, T., Pijnenburg, Y. A., van der Flier, W. M., ... & Ramakers, I. (2013). Progression to dementia in memory clinic patients without dementia A latent profile analysis. *Neurology*, *81*(15), 1342-1349. doi:10.1212/WNL.0b013e3182a82536

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural equation modeling: a multidisciplinary journal*, *20*(1),1-26. doi:10.1080/10705511.2013.742377

Lezak, M. D. (2004). *Neuropsychological assessment*. New York, NY: Oxford University Press.

Libon, D. J., Drabick, D. A., Giovannetti, T., Price, C. C., Bondi, M. W., Eppig, J., ... & Nation, D. A. (2014). Neuropsychological syndromes associated with Alzheimer's/vascular dementia: a latent class analysis. *Journal of Alzheimer's Disease*, *42*(3), 999-1014. doi:10.3233/JAD-132147

Libon, D. J., Xie, S. X., Eppig, J., Wicas, G., Lamar, M., Lippa, C., ... & Wambach, D. M. (2010). The heterogeneity of mild cognitive impairment: A neuropsychological analysis. *Journal of the International Neuropsychological Society*, *16*(01), 84-93. doi:10.1017/S1355617709990993

Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, *20*(1), 36-43. doi:10.1.1.128.9157

Mapstone, M., Steffenella, T. M., & Duffy, C. J. (2003). A visuospatial variant of mild cognitive impairment Getting lost between aging and AD. *Neurology*, *60*(5), 802-808. doi:10.1212/01.WNL.0000049471.76799.DE

McGuinness, B., Barrett, S. L., McIlvenna, J., Passmore, A. P., & Shorter, G. W. (2015). Predicting conversion to dementia in a memory clinic: A standard clinical approach compared with an empirically defined clustering method (latent profile analysis) for mild cognitive impairment subtyping. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *1*(4), 447-454. doi:10.1016/j.dadm.2015.10.003

McKeith, I. G., Galasko, D., Kosaka, K., Perry, E. K., Dickson, D. W., Hansen, L. A., ... & Lennox, G. (1996). Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB) Report of the consortium on DLB international workshop. *Neurology*, *47*(5), 1113-1124. doi:10.1212/WNL.47.5.1113

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

Molano, J., Boeve, B., Ferman, T., Smith, G., Parisi, J., Dickson, D., ... & Kantarci, K. (2010). Mild cognitive impairment associated with limbic and neocortical Lewy body disease: a clinicopathological study. *Brain*, *133*(2), 540-556. doi:10.1093/brain/awp280

Morin, R. T., & Axelrod, B. N. (2017). Use of Latent Class Analysis to define groups based on validity, cognition, and emotional functioning. *The Clinical Neuropsychologist*, 1-13. doi:10.1080/13854046.2017.1341550

Muthén, B. (2004). Latent variable analysis. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences (*pp. 345-368). Thousand Oaks, CA: Sage Publications.

Muthén, L. K. & Muthén, B. O. (2006). Chi-square difference testing using the Satorra-Bentler scaled chi-square. *Online technical appendix.* Los Angeles, CA: Muthén & Muthén. Retrieved from https://www.statmodel.com/chidiff.shtml

Muthén, L. K., & Muthén, B. O. (2012). *Mplus Version 7 user’s guide.* Los Angeles, CA: Muthén & Muthén.

Nielson, K. A., Cummings, B. J., & Cotman, C. W. (1996). Constructional apraxia in Alzheimer's disease correlates with neuritic neuropathology in occipital cortex. *Brain research*, *741*(1), 284-293. doi: 10.1016/S0006-8993(96)00983-3

Norman, M. A., Moore, D. J., Taylor, M., Franklin Jr, D., Cysique, L., Ake, C., ... & HNRC Group. (2011). Demographically corrected norms for African Americans and Caucasians on the hopkins verbal learning test–revised, brief visuospatial memory test–revised, stroop color and word test, and wisconsin card sorting test 64-card version. *Journal of clinical and experimental neuropsychology*, *33*(7), 793-804. doi:10.1080/13803395.2011.559157

Nylund, K. L. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization* (Doctoral dissertation, University of California, Los Angeles). Retrieved from http://www.statmodel.com/download/Nylund%20dissertation%20Updated1.pdf

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, *14*(4), 535-569. doi:10.1080/10705510701575396

Peraita, H., Chacón, J., Díaz-Mardomingo, C., & Martínez-Arias, R. (2015). Profiles of mild cognitive impairment (MCI) in the elderly. *The Spanish journal of psychology*, *18,* 1-12. doi:10.1017/sjp.2015.96

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

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., ... & Trojanowski, J. Q. (2010). Alzheimer's disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*, *74*(3), 201-209. doi:10.1212/WNL.0b013e3181cb3e25

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., ... & Winblad, B. (2001). Current concepts in mild cognitive impairment. *Archives of neurology*, *58*(12), 1985-1992. doi:10.1001/archneur.58.12.1985

Petersen, R. C., & Morris, J. C. (2005). Mild cognitive impairment as a clinical entity and treatment target. *Archives of neurology*, *62*(7), 1160-1163. doi:10.1001/archneur.62.7.1160

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. Archives of neurology, 56(3), 303-308. doi:10.1017/S1041610204000067

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International journal of behavioral development*, *33*(6), 565-576. doi:10.1177/0165025409343765

Reisberg, B., Ferris, S. H., Kluger, A., Franssen, E., Wegiel, J., & De Leon, M. J. (2008). Mild cognitive impairment (MCI): a historical perspective. *International Psychogeriatrics*, *20*(1), 18-31. doi:10.1017/S1041610207006394

Roesch, S. C., Villodas, M., & Villodas, F. (2010). Latent class/profile analysis in maltreatment research: A commentary on Nooner et al., Pears et al., and looking beyond. *Child abuse & neglect*, *34*(3), 155-160. doi:10.1016/j.chiabu.2010.01.003

Satorra, A. & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243-248. doi:10.1007/S11336-009-9135-Y

Saunders, A. M., Strittmatter, W. J., Schmechel, D., George-Hyslop, P. S., Pericak-Vance, M. A., Joo, S. H., ... & Hulette, C. (1993). Association of apolipoprotein E allele ε4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*, *43*(8), 1467-1467. doi:10.1212/WNL.43.8.1467

Schindler, S. E., Gray, J. D., Gordon, B. A., Xiong, C., Batrla-Utermann, R., Quan, M., Wahl, S., Benzinger, T. L. S., Holtzman, D. M., Morris, J. C., & Fagan, A. M. (2018). Cerebrospinal fluid biomarkers measured by Elecsys assays compared to amyloid imaging. *Alzheimer's & Dementia*, *14*(11), 1460-1469. doi:10.1016/j.jalz.2018.01.013

Schneider, J. A., Arvanitakis, Z., Leurgans, S. E., & Bennett, D. (2009). The neuropathology of probable Alzheimer disease and mild cognitive impairment. *Annals of neurology*, *66*(2), 200-208. doi: 10.1002/ana.21706

Shao, A., Liang, L., Yuan, C., & Bian, Y. (2014). A latent class analysis of bullies, victims and aggressive victims in chinese adolescence: relations with social and school adjustments. *PloS one*, *9*(4), e95290. doi:10.1371/journal.pone.0095290

Shaw, L. M., Vanderstichele, H., Knapik-Czajka, M., Clark, C. M., Aisen, P. S., Petersen, R. C., ... & Trojanowski, J. Q. (2009). Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of neurology*, *65*(4), 403-413. doi:10.1002/ana.21610

Sperling, R. A., Rentz, D. M., Johnson, K. A., Karlawish, J., Donohue, M., Salmon, D. P., & Aisen, P. (2014). The A4 study: stopping AD before symptoms begin?. *Science translational medicine*, *6*(228), 228fs13-228fs13. doi:10.1126/scitranslmed.3007941

Stein, J., Luppa, M., Brähler, E., König, H. H., & Riedel-Heller, S. G. (2010). The assessment of changes in cognitive functioning: Reliable change indices for neuropsychological instruments in the elderly–a systematic review. *Dementia and geriatric cognitive disorders*, *29*(3), 275-286. doi:10.1159/000289779

Tein, J. Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural equation modeling: a multidisciplinary journal*, *20*(4), 640-657. doi:10.1080/10705511.2013.824781

Teng, E., Becker, B. W., Woo, E., Knopman, D. S., Cummings, J. L., & Lu, P. H. (2010). Utility of the Functional Activities Questionnaire for distinguishing mild cognitive impairment from very mild Alzheimer's disease. *Alzheimer disease and associated disorders*, *24*(4), 348. doi:10.1097/WAD.0b013e3181e2fc84

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley Publishing Company.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., ... & Morris, J. C. (2013). The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & Dementia*, *9*(5), e111-e194. doi:10.1016/j.jalz.2013.05.1769

Wilson, R. S., Yu, L., Trojanowski, J. Q., Chen, E. Y., Boyle, P. A., Bennett, D. A., & Schneider, J. A. (2013). TDP-43 pathology, cognitive decline, and dementia in old age. *JAMA neurology*, *70*(11), 1418-1424. doi:10.1001/jamaneurol.2013.3961

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., ... & Arai, H. (2004). Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of internal medicine*, *256*(3), 240-246. doi:10.1111/j.1365-2796.2004.01380.x

Zlokovic, B. V. (2011). Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nature Reviews Neuroscience*, *12*(12), 723-738. doi:10.1038/nrn3114