

Electronic Thesis and Dissertation Repository

8-18-2017 11:00 AM

Cognition and Motor Function: A Novel Outcome Measure for Studies on Pre-Dementia Syndromes

Jacqueline K. Kueper, *The University of Western Ontario*

Supervisor: Drs. Mark Speechley, *The University of Western Ontario*

Joint Supervisor: Manuel Montero-Odasso, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Epidemiology and Biostatistics

© Jacqueline K. Kueper 2017

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Clinical Epidemiology Commons](#)

Recommended Citation

Kueper, Jacqueline K., "Cognition and Motor Function: A Novel Outcome Measure for Studies on Pre-Dementia Syndromes" (2017). *Electronic Thesis and Dissertation Repository*. 4776.

<https://ir.lib.uwo.ca/etd/4776>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Advances in dementia research have shifted attention towards earlier stages in the natural history, such as Mild Cognitive Impairment. The current gold standard outcome measure, the Alzheimer's Disease Assessment Scale-Cognitive Subscale, is not optimally responsive to changes in pre-dementia populations. Modifications to scoring methodology and content have improved the measurement performance of the ADAS-Cog. However, no published modifications have addressed a second key shift in the field towards understanding motor function as an important component of dementia and pre-dementia syndromes. This thesis used a Pooled Index approach to combine an ADAS-Cog-Proxy measure with assessments of gait velocity and dual-task cost. The responsiveness of the PI to baseline discrimination between older adults with normal cognition, Subjective Cognitive Impairment, and MCI was similar to the ADAS-Cog-Proxy. The PI demonstrated greater responsiveness than the ADAS-Cog-Proxy to change over 6mo. and 48mo., but not 36mo. of follow-up. Overall, motor function assessments improve ADAS-Cog responsiveness.

Keywords

Mild Cognitive Impairment, Dementia, Outcome Measurement, Alzheimer's Disease Assessment Scale-Cognitive Subscale, Motor Function, Gait, Dual-Task Cost

Co-Authorship Statement

Jacqueline K. Kueper (JKK) wrote all components of this thesis and is the primary author. JKK was assisted by Dr. Mark Speechley (MS), Dr. Manuel Montero-Odasso (MMO), and Dr. Daniel J. Lizotte (DJL). In accordance with the authorship criteria recommended by the International Committee of Medical Journal Editors,¹ contributions were as follows:

1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work (JKK,MS,MMO,DJL);
2. Drafting the work (JKK) or revising it critically for important intellectual content (MS,MMO,DJL);
3. Final approval of the version to be published (JKK,MS,MMO,DJL);
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved (JKK,MS,MMO,DJL).

In addition, some of the data used in the preparation of this thesis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Reference:

1. International Committee of Medical Journal Editors. Defining the role of authors and contributors. <http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>. Published 2017. Accessed July 14, 2017.

Acknowledgments

I would first like to acknowledge the Anishinaabeg, Haudenosaunee, Leni-Lunaape, and Attawandaron (Neutral) peoples, upon whose traditional territory the majority of this thesis was conceptualized and created.

I would next like to thank my supervisors, Drs. Mark Speechley and Manuel Montero-Odasso, for their guidance and encouragement towards this thesis and throughout other aspects of the master's program. I am grateful for the opportunities for growth and development that they have provided me with.

I would like to thank Dr. Daniel J. Lizotte for joining my advisory committee, and for his unwavering support and insights.

I would like to acknowledge the members of the Gait and Brain Lab for all the work that they do, and thank them for their infectious energy of enthusiasm and purpose. In particular, thank you to Navena Lingum, Alanna Black, Yanina Sarquis Adamson, Frederico Faria, Korbin Blue, and Brittany Barnes.

Randy Zabukovec, Doug Freeman, Kurtis Marlow, and Dr. Monica Volesky have each made widespread contributions to my wellbeing and personal walking ability. Although not necessarily obvious, they have been instrumental to the completion of this thesis.

The Alzheimer Society of London and Middlesex generously provided funding towards this thesis.

I would also like to acknowledge the Alzheimer's Disease Neuroimaging Initiative, which permitted me to use their data. Thus, data collection and sharing for a portion of this thesis was funded by ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute of Biomedical Imaging and Bioengineering, and through contributions from: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company;

Eurolmmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merk & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The other source of data for this thesis, the Gait and Brain Study, is supported by a CIHR Operating Grant (MOP 211220) and a CIHR Project Grant (PJT 153100). Additional funding has come from the Early Research Ontario Award (PI Dr. Montero-Odasso).

Last but certainly not least, I would like to express my gratitude towards the community of family, friends, and fellow students (not mutually exclusive categories) that I am fortunate to be a part of. Thank you, for everything.

Table of Contents

| | |
|---|-----|
| Abstract | i |
| Co-Authorship Statement..... | ii |
| Acknowledgments..... | iii |
| Table of Contents | v |
| List of Tables | x |
| List of Figures | xii |
| List of Abbreviations | xiv |
| Chapter 1 | 1 |
| 1 Introduction | 1 |
| 1.1 Epidemiology of Dementia | 1 |
| 1.2 Outcome Measurement Challenges | 1 |
| 1.3 Outcome Measurement Challenges in Dementia Research | 2 |
| 1.4 Overview of Thesis | 4 |
| 1.5 References | 5 |
| Chapter 2 | 8 |
| 2 Introduction to Outcome Measures | 8 |
| 2.1 Overview of Latent Traits | 8 |
| 2.2 Classical Test Theory..... | 11 |
| 2.2.1 Standardization | 12 |
| 2.2.2 Reliability..... | 13 |
| 2.2.3 Precision..... | 14 |
| 2.2.4 Standard Error of Measurement..... | 14 |
| 2.2.5 Coefficient of Variation | 14 |
| 2.2.6 Validity | 15 |

| | | |
|-----------|---|----|
| 2.2.7 | Responsiveness | 17 |
| 2.2.8 | Sensitivity and Specificity | 20 |
| 2.3 | Item Response Theory | 21 |
| 2.4 | Approaches for Modifying Outcome Measures..... | 24 |
| 2.4.1 | Statistical Modification of Scoring Methodology..... | 25 |
| 2.4.2 | Adding Additional Item Content | 25 |
| 2.5 | Key Cognitive Outcome Measures | 26 |
| 2.6 | Summary | 27 |
| 2.7 | References..... | 28 |
| Chapter 3 | | 31 |
| 3 | Literature Review..... | 31 |
| 3.1 | Development of the Alzheimer’s Disease Assessment Scale | 31 |
| 3.1.1 | Item Selection | 31 |
| 3.1.2 | Validation of the Alzheimer’s Disease Assessment Scale..... | 34 |
| 3.1.3 | Responsiveness to Baseline Discrimination | 35 |
| 3.1.4 | Responsiveness to Disease Progression..... | 36 |
| 3.1.5 | Minimum Clinically Relevant Change | 37 |
| 3.1.6 | Summary | 37 |
| 3.2 | Assessment of the ADAS-Cog 11 in Pre-Dementia Populations | 37 |
| 3.2.1 | Ceiling Effects | 37 |
| 3.2.2 | Information Content..... | 38 |
| 3.2.3 | Invariances | 39 |
| 3.2.4 | Reliability..... | 40 |
| 3.2.5 | Concurrent Criterion Validation | 40 |
| 3.2.6 | Responsiveness at the Item Level | 40 |

| | | |
|--------|--|----|
| 3.2.7 | Performance of the ADAS-Cog 11 as an Outcome Measure in Pre-Dementia Studies | 41 |
| 3.2.8 | Summary of ADAS-Cog 11 Performance in Pre-Dementia Studies | 60 |
| 3.3 | Modifications of the ADAS-Cog 11 | 60 |
| 3.3.1 | ADAS-Cog 13..... | 61 |
| 3.3.2 | Vascular Dementia Assessment Scale | 63 |
| 3.3.3 | ADAS-Cog 12..... | 64 |
| 3.3.4 | Test for the Early Detection of Dementia from Depression – Cognitive.. | 67 |
| 3.3.5 | Pooled Index | 68 |
| 3.3.6 | ADAS-Rasch..... | 69 |
| 3.3.7 | ADAS-Tree | 71 |
| 3.3.8 | Computerized ADAS-Cog | 73 |
| 3.3.9 | Touch Panel-Type Dementia Assessment Scale..... | 74 |
| 3.3.10 | Computerized Adaptive Testing of the Cambridge Cognitive Examination – Plus | 75 |
| 3.3.11 | ADAS-Cog-5-Subset and ADAS-Cog-6-Subset | 76 |
| 3.3.12 | The ADAS-Cog-Plus (ADAS-Bifactor, ADAS-Plus-EF, ADAS-Plus-EF&FA) | 77 |
| 3.3.13 | Common Item Pooling..... | 79 |
| 3.3.14 | Alzheimer’s Disease Neuroimaging Initiative Memory Composite..... | 80 |
| 3.3.15 | ADAS-Cog IRT | 81 |
| 3.3.16 | ADAS-3 | 83 |
| 3.3.17 | Cognitive Composites 1 and 2 | 84 |
| 3.3.18 | Cognitive Functional Composites 1 and 2 | 85 |
| 3.3.19 | Item Response Theory and Pharmacometric ADAS-Cog 13 | 87 |
| 3.3.20 | integrated Alzheimer’s Disease Rating Scale | 90 |
| 3.3.21 | Straightforward Sensitive Scale | 91 |

| | | |
|-----------|--|-----|
| 3.3.22 | ADAS-Cog 3b..... | 92 |
| 3.3.23 | ADAS-Cog 5..... | 93 |
| 3.3.24 | ADAS-13 Re-Weighted | 94 |
| 3.3.25 | Alzheimer’s Disease Composite Score | 95 |
| 3.3.26 | Summary of Modifications Made to the ADAS-Cog 11 | 96 |
| 3.4 | Summary of Chapter 3 | 97 |
| 3.5 | References | 98 |
| Chapter 4 | | 104 |
| 4 | Research Question and Objectives..... | 104 |
| 4.1 | Research Question | 104 |
| 4.2 | Objectives | 104 |
| 4.2.1 | Objective 1 | 104 |
| 4.2.2 | Objective 2 | 104 |
| 4.2.3 | Objective 3 | 105 |
| 4.3 | Conclusion | 105 |
| Chapter 5 | | 106 |
| 5 | Integrated Article: Cognition and Motor Function: The Gait and Brain Pooled Index..... | 106 |
| 5.1 | Introduction..... | 106 |
| 5.2 | Methods..... | 109 |
| 5.2.1 | Study Population..... | 109 |
| 5.2.2 | Measures | 110 |
| 5.2.3 | Data Analysis | 113 |
| 5.3 | Results..... | 115 |
| 5.3.1 | Baseline Characteristics | 115 |
| 5.3.2 | ADAS-Cog Proxy Model..... | 118 |

| | | |
|------------------|--|-----|
| 5.3.3 | Gait and Brain Pooled Index..... | 118 |
| 5.3.4 | Baseline Discrimination..... | 118 |
| 5.3.5 | Change Over Time..... | 119 |
| 5.4 | Discussion..... | 120 |
| 5.5 | References..... | 122 |
| Chapter 6 | | 125 |
| 6 | Detailed Methods..... | 125 |
| 6.1 | Objective 1..... | 125 |
| 6.1.1 | Step 1: Develop an Alzheimer’s Disease Assessment Scale – Cognitive Subscale Proxy..... | 125 |
| 6.1.2 | Step 2: Select Additional Measures for the Novel Outcome Measure ... | 137 |
| 6.1.3 | Step 3: Combine Measures Using a Pooled Index Approach..... | 145 |
| 6.2 | Objective 2..... | 146 |
| 6.3 | Objective 3..... | 148 |
| 6.4 | References..... | 150 |
| Chapter 7 | | 151 |
| 7 | Discussion..... | 151 |
| 7.1 | Summary of Findings..... | 151 |
| 7.2 | Comparison with other ADAS-Cog Modifications..... | 154 |
| 7.3 | Limitations..... | 157 |
| 7.4 | Suggestions for Further Research..... | 159 |
| 7.5 | Clinical Implications..... | 161 |
| 7.6 | Conclusion..... | 162 |
| 7.7 | References..... | 163 |
| Appendices | | 167 |
| Curriculum Vitae | | 178 |

List of Tables

| | |
|---|-----|
| Table 1 ADAS-Cog 11 Ceiling Effects. | 38 |
| Table 2 Normal Cognition Samples: ADAS-Cog 11 responsiveness to group-level between-person differences in observed level of disease severity based on exposure status. | 42 |
| Table 3 Mild Cognitive Impairment Samples: ADAS-Cog 11 responsiveness to group-level between-person differences in observed level of disease severity based on exposure status | 43 |
| Table 4 Mixed Normal Cognition and Mild Cognitive Impairment Samples: ADAS-Cog 11 responsiveness to group-level between-person differences in observed level of disease severity based on exposure status. | 47 |
| Table 5 Normal Cognition Samples: Responsiveness to group-level between-person differences of within-person observed change in those estimated to be different based on baseline exposure status. | 48 |
| Table 6 Mild Cognitive Impairment Samples: Responsiveness to group-level between-person differences of within-person observed change in those estimated to be different based on baseline exposure status. | 51 |
| Table 7 ADAS-Cog 11 Responsiveness to Treatment Effects in Pre-Dementia Clinical Trials | 55 |
| Table 8 Sample Size Estimates to Detect Treatment Effects in Pre-Dementia Clinical Trials | 58 |
| Table 9 Alzheimer's Disease Neuroimaging Initiative Baseline Characteristics | 115 |
| Table 10 Gait and Brain Study Baseline Characteristics | 116 |
| Table 11 Standardized Response Means: Responsiveness to group-level within-person measured change over time..... | 119 |

| | |
|---|-----|
| Table 12 Cognitive Tests in Available Databases at Baseline..... | 126 |
| Table 13 Range of ADAS-Cog-Proxy Covariate Scores in ADNI and GABS | 133 |
| Table 14 Missing ADAS-Cog-Proxy Covariates in the Gait and Brain Study..... | 134 |
| Table 15 Correlation Coefficients for Potential Predictor Matrix Variables | 135 |
| Table 16 ADAS-Cog-Proxy Scores in the Gait and Brain Study | 137 |
| Table 17 Correlation Coefficients for Quantitative Gait Parameters and ADAS-Cog-Proxy | 138 |
| Table 18 Correlation Coefficients for Dual-Task Cost Assessments and ADAS-Cog-Proxy | 138 |
| Table 19 Correlation Coefficients for Single Task and Dual-Task Cost Gait Variables that were Retained after Steps One to Three | 139 |
| Table 20 Pairwise Correlation Coefficients between Final Pooled Index Components.. | 145 |

List of Figures

| | |
|--|-----|
| Figure 1 Categorical latent trait. | 9 |
| Figure 2 Dimensional latent trait. | 10 |
| Figure 3 Factor mixed model latent trait..... | 10 |
| Figure 4 Unidimensionality assumption. | 11 |
| Figure 5 Sample item characteristic curve..... | 22 |
| Figure 6 Timeline of ADAS-Cog 11 modifications. | 61 |
| Figure 7 ADAS-Cog-Proxy model development and application. | 111 |
| Figure 8 Residuals between observed and predicted ADAS-Cog scores in the ADNI testing data. | 131 |
| Figure 9 ADAS-Cog-Proxy model. | 132 |
| Figure 10 Box plot of gait velocity by baseline diagnostic category..... | 140 |
| Figure 11 Box plot of gait stride time by baseline diagnostic category..... | 140 |
| Figure 12 Box plot of gait stride time variability by baseline diagnostic category. | 141 |
| Figure 13 Box plot of dual-task gait velocity cost when counting backwards by ones against baseline diagnostic category..... | 141 |
| Figure 14 Box plot of dual-task gait velocity cost when counting backwards by serial sevens against baseline diagnostic category. | 142 |
| Figure 15 Box plot dual-task gait velocity cost when naming animals against baseline diagnostic category. | 142 |
| Figure 16 Box plot of dual-task stride time variability cost when naming animals against baseline diagnostic category. | 143 |

Figure 17 Scatterplot of gait velocity against ADAS-Cog-Proxy scores. 144

Figure 18 Scatterplot of dual-task gait velocity cost with secondary task of counting backwards by ones against ADAS-Cog-Proxy scores. 144

Figure 19 Scatterplot of dual-task gait velocity cost with secondary task of counting backwards by ones against gait velocity. 145

Figure 20 Box plot of a pooled index that includes an assessment of instrumental activities of daily living against baseline diagnostic category. 147

Figure 21 Box plot of pooled index against baseline diagnostic category. 147

List of Abbreviations

AD: Alzheimer's Disease

ADAS: Alzheimer's Disease Assessment Scale

ADAS-Cog: Alzheimer's Disease Assessment Scale – Cognitive Subscale

ADAS-Cog 11: Original 11 item version of the Alzheimer's Disease Assessment Scale – Cognitive Subscale

ADAS-Cog-Proxy: Generalized Additive Model that can be used to approximate Alzheimer's Disease Assessment Scale – Cognitive Subscale scores

ADAS-Cog 3/3b/5/5-Subset/6-Subset/9/12/13/13RW/14/Bifactor/IRT/Plus-EF/Plus-EF&FA/Rasch/Tree: Modified versions of the Alzheimer's Disease Assessment Scale – Cognitive Subscale (See Chapter 3 for descriptions)

ADAS-Noncog: Alzheimer's Disease Assessment Scale – Non-Cognitive Subscale

ADCOMS: Alzheimer's Disease Composite Score

ADL: Activities of Daily Living

ADNI: Alzheimer's Disease Neuroimaging Initiative

ADNI-Mem: Alzheimer's Disease Neuroimaging Initiative Memory Composite

AI: Apathy Inventory

aMCI: Amnesic Mild Cognitive Impairment

ANART: American National Adult Reading Test

APOEε4: Apolipoprotein E gene, E4 allelic form

AUC: Area Under the Curve

A β : Beta-amyloid deposition

BADS: Behavioural Assessment of the Dysexecutive Syndrome

BMI: Body Mass Index

BNT: Boston Naming Test

BVRT: Benton Visual Retention Test

cADAS-Cog: Computerized Alzheimer's Disease Assessment Scale – Cognitive Subscale

CAMCOG: Cambridge Cognition Examination

CAT: Computerized Adaptive Testing

CC: Cognitive Composite

CDR: Clinical Dementia Rating Scale

CDR-SB: Clinical Dementia Rating Scale – Sum of Boxes

CFC: Cognitive Functional Composite

CHD: Coronary Heart Disease

ChEI: Cholinesterase Inhibitors

CI: Confidence Interval

CIBIC-Plus: Clinician's Interview-Based Impression of Change plus Caregiver Input

cm/s: Centimeters per second

CMINDS: Computerized Multiphasic Interactive Neurocognitive Dual Display System

COWAT: Controlled Oral Words Association Test

CSF: Cerebrospinal Fluid

CTT: Classical Test Theory

CV: Coefficient of Variation

CVD: Cardiovascular Disease

CVLT: California Verbal Learning Test

DBRI: Dysfunctional Behaviour Rating Instrument

DIF: Differential Item Functioning

DSfwd/bkwd: Digit Span Forwards/Backwards Test

DSS: Digit Symbol Substitution

DTC: Dual-Task Cost

EF: Executive Function

FAQ: Functional Activities Questionnaire

GAB-PI: Gait and Brain Pooled Index

GABS: Gait and Brain Study

GAM: Generalized Additive Model

IADL: Instrumental Activities of Daily Living

iADRS: Integrated Alzheimer's Disease Rating Scale

ICC: Item Characteristic Curve

IRT: Item Response Theory

MADRS: Montgomery-Asberg Depression Rating Scale

MCI: Mild Cognitive Impairment

MICE: Multiple Imputation by Chained Equations

MMSE: Mini-Mental State Examination

Mo.: Months

MoCA: Montreal Cognitive Assessment

ms: milliseconds

n: sample size

NC: Normal Cognition

OR: Odds Ratio

pG/mL: picograms per milliliter

PRE: Perception Response Evaluation

p-Tau: Phosphorylated Tau protein

PET: Positron Emission Tomography

PI: Pooled Index

QoL: Quality of Life

RAVLT: Rey Auditory Verbal Learning Test

RBANS: Repeatable Battery for the Assessment of Neuropsychological Status

RF: Random Forests

ROC: Receiver Operating Characteristics

SCI: Subjective Cognitive Impairment

SCOPA-Cog: Scales for Outcome of Parkinson's Disease-Cognition

SD: Standard Deviation

SDMT: Symbol Digit Modalities Test

SEM: Standard Error of Measurement

SRM: Standardized Response Mean

SRT: Selective Reminding Test

SSS: Straightforward Sensitive Scale

SUVr: Standard Uptake Values Relative to Cerebellum

SYNERGIC: Synchronizing Exercises, Remedies in Gait and Cognition

TDAS: Touch Panel – Type Dementia Assessment Scale

TE4D-Cog: Test for the Early Detection of Dementia from Depression

TMT A&B: Trail Making Test, forms A and B

t-Tau: Total Tau protein

VaD: Vascular Dementia

VaDAS: Vascular Dementia Assessment Scale

Vs.: Versus

WAIS: Weschler Adult Intelligence Scale

WMS: Wechsler Memory Scale

3MSE: Modified Mini-Mental State Examination

Chapter 1

1 Introduction

The purpose of Chapter 1 is to provide a brief overview of this thesis. Three outcome measurement challenges will be addressed, which are currently present in the field of dementia research.

1.1 Epidemiology of Dementia

Dementia is a syndrome characterized by deterioration in cognitive abilities such as memory, praxis, and language, and in the ability to perform everyday activities.^{1,2} The worldwide prevalence of dementia is 47 million people, with an estimated incidence rate of 9.9 million cases per year.¹ In 2016, an estimated 564,000 people living in Canada had dementia, costing an annual \$10.4 billion.³ The prevalence of dementia in Canada is expected to reach 912,000 cases by 2030.³ There is no known cure. Hence, much research is aimed at trying to better understand dementia syndromes and develop effective treatment approaches.

1.2 Outcome Measurement Challenges

The quality of any research study is influenced by the measurement tools employed to assess constructs of interest.^{4,5} Because a construct is a hypothetical concept, a fundamental challenge lies in valid and reliable measurement.⁶ In the context of health research, constructs are often aspects of disease pathology or encapsulate the impacts that pathology may have on one's experience of life; they are dynamic yet bounded by the current understanding of a health condition or state. In some cases, a 'gold standard' or best possible outcome measure has been established. Beyond individual study quality, gold standards help to increase consistency and comparability throughout a body of literature, which is especially important when evaluating novel treatment approaches. However, as a field advances the understanding of a health condition or state, including what constitutes pathology or burden, and ultimately treatment benefit, may change. If a

gold standard is not harmonious with these advancements, the quality and relevance of research findings, and by extension the speed with which a field progresses, may be limited.

The circumstance of a long-standing gold standard in a rapidly advancing field is the first of three challenges pertaining to outcome measurement that this thesis will address. The second is how an outcome measure may be modified for improvement if the original version is deemed unsatisfactory for use in a particular population or context. The third is when all necessary outcome measures for a research objective are not available in a single database, but a preliminary test of hypotheses is desired before investing the time and resources required to run a new study that would collect all measures together.

1.3 Outcome Measurement Challenges in Dementia Research

These three measurement challenges will be examined in the field of dementia research, where the gold standard for assessing the efficacy of antidementia therapies is the Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog).⁷ Although the ADAS-Cog works well for studies on dementia populations, the research field is changing in such a way that the ADAS-Cog is now being used in contexts that it was not originally developed for.

More specifically, two important advancements in the field of dementia research have occurred since the adoption of the ADAS-Cog in the 1980s. First, research interest has shifted to pre-dementia stages of disease progression, such as Mild Cognitive Impairment (MCI), where impairment is more mild than in dementia.⁸⁻¹³ It is thought that intervening to slow or stop the progression of disease will be more effective than waiting until severe neuropathology and dysfunction have developed.^{8,11,14-16} Thus, many research studies, both observational and experimental, are being conducted in pre-dementia populations.^{12,15,17-21} Outcome measures that work well for studies of older adults with dementia may not work well for studies of older adults with pre-dementia syndromes because the impairment that occurs in pre-dementia syndromes is more mild than the

impairment that occurs in dementia syndromes, and may differ in type. For example, memory is often impaired at early stages while language does not become impaired until more severe stages.²² It is important that outcome measures being used for pre-dementia populations can reflect a person's level of cognitive ability, as well as changes in disease severity over time, otherwise disease progression or potential treatment benefits may be missed. Concerns have been raised about whether the ADAS-Cog, which was originally developed to assess dysfunction in mild to severe dementia, is able to detect important changes at earlier stages of disease progression.^{9,14,23,24} These concerns relate to the first measurement challenge introduced above, and motivated a literature review for this thesis that explores the measurement properties and performance of the ADAS-Cog in pre-dementia populations. In accordance with the second measurement challenge, the review extends to document all modifications that have been made in an attempt to improve the ADAS-Cog. This literature review is presented in Chapter 3.

The second advancement in the field of dementia research is the emergence of motor function decline as an early pathological manifestation, in addition to cognitive decline, of disease progression; at the time of ADAS-Cog development, cognitive and motor function decline were understood as separate processes. A seminal study in 1997 found older adults who stop walking while talking are at an increased risk of falls compared to those who do not stop.²⁵ Since then, a literature base has been growing that supports an association of dementia and pre-dementia syndromes with both motor and cognitive decline, whereby these declines are understood as interrelated processes.²⁶⁻⁴¹ For example, Buracchio et al. (2010) found walking speed begins to slow twelve years in advance of MCI diagnosis,²⁶ Montero-Odasso et al. (2014) suggest that subtypes of MCI possess a unique "motor signature",²⁷ and Kueper et al. (2017) performed a systematic review that found poor lower limb motor performance is associated with an increased risk of incident dementia.²⁸ Importantly, motor function and cognitive abilities together may provide the fundamental basis for functionality, or the ability to perform activities of daily living, the loss of which is a hallmark of disease severity.⁴²⁻⁴⁴ Motor function assessments may therefore be helpful for detecting important changes in pre-dementia syndromes, such as to evaluate whether a novel treatment approach is beneficial. However, the literature review on the ADAS-Cog did not find any modifications that incorporate motor

function assessments. The main question this thesis aims to address is whether adding assessments of motor function to the ADAS-Cog improves its ability to detect changes in pre-dementia syndromes. This research question is presented in Chapter 4 along with three formal objectives, which include developing an outcome measure and assessing its ability to detect two types of change in a pre-dementia sample. Chapter 5 includes a version of a manuscript centered around these objectives.

The third measurement challenge presented above becomes relevant as this thesis relies on secondary data analysis, and no database contains both the ADAS-Cog and motor function assessments. A proxy ADAS-Cog was developed for use in a database that contains motor function assessments. The framework used to build this proxy ADAS-Cog may be followed for other, similar situations, and is covered in depth in Chapter 6 along with other detailed methods and results pertaining to the three objectives.

1.4 Overview of Thesis

The next chapter provides an introduction to outcome measurement terms and concepts that will be utilized throughout the remainder of the thesis, Chapter 3 presents a literature review on the ADAS-Cog, Chapter 4 states the research question and objectives, Chapter 5 is an integrated article, Chapter 6 includes more detailed methods and results than are presented in Chapter 5, Chapter 7 provides an extended discussion, and the Appendices contain supplementary Tables and Figures.

1.5 References

1. World Health Organization. Dementia. <http://www.who.int/mediacentre/factsheets/fs362/en/>. Published 2017. Accessed July 2, 2017.
2. Alzheimer Society of Canada. What is dementia? <http://www.alzheimer.ca/en/About-dementia/What-is-dementia>. Published 2017. Accessed July 2, 2017.
3. Chambers LW, Bancej C, McDowell I. *Prevalence and Monetary Costs of Dementia in Canada*. Toronto: Alzheimer Society of Canada; 2016.
4. Coster WJ. Making the best match: selecting outcome measures for clinical trials and outcome studies. *Am J Occup Ther*. 2013;67(2):162-170. doi:10.5014/ajot.2013.006015.
5. Fullerton JT. Evaluation of research studies Part IV: validity and reliability-concepts and application. *J Nurse Midwifery*. 1993;38(2):121-125.
6. Binning JF. Construct. In: *Encyclopedia Britannica Online*. Encyclopedia Britannica Inc. 2017. <https://www.britannica.com/science/construct>.
7. Rosen WG, Mohs RC, Davis KL. A New Rating Scale for Alzheimer's Disease. *Am J Psychiatry*. 1984;141(11):1356-1364. doi:10.1176/ajp.141.11.1356.
8. Llano DA, Laforet G, Devanarayan V. Derivation of a new ADAS-Cog composite using tree-based multivariate analysis: prediction of conversion from Mild Cognitive Impairment to Alzheimer Disease. *Alzheimer Dis Assoc Disord*. 2011;25(1):73-84. doi:10.1097/WAD.0b013e3181f5b8d8.
9. Podhorna J, Krahnke T, Shear M, E Harrison J. Alzheimer's Disease Assessment Scale-Cognitive subscale variants in Mild Cognitive Impairment and mild Alzheimer's Disease: change over time and the effect of enrichment strategies. *Alzheimers Res Ther*. 2016;8(1):8. doi:10.1186/s13195-016-0170-5.
10. Graham DP, Cully JA, Snow AL, Massman P, Doody R. The Alzheimer's Disease Assessment Scale-Cognitive subscale: normative data for older adult controls. *Alzheimer Dis Assoc Disord*. 2004;18(4):236-240.
11. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Kokmen E, Tangelos EG. Aging, memory, and Mild Cognitive Impairment. *Int psychogeriatrics*. 1997;9(Suppl 1):65-69.
12. Cooper C, Sommerlad A, Lyketsos CG, Livingston G. Modifiable predictors of dementia in Mild Cognitive Impairment: a systematic review and meta-analysis. *Am J Psychiatry*. 2015;172(4):323-334. doi:10.1176/appi.ajp.2014.14070878.
13. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild Cognitive Impairment: a concept in evolution. *J Intern Med*. 2014;275(3):214-228. doi:10.1111/joim.12190.
14. Ueckert S, Plan EL, Ito K, Karlsson MO, Corrigan B, Hooker AC. Improved utilization of ADAS-Cog assessment data through Item Response Theory based Pharmacometric Modeling. *Pharm Res*. 2014;31(8):2152-2165. doi:10.1007/s11095-014-1315-5.
15. Mariani E, Monastero R, Mecocci P. Mild Cognitive Impairment: a systematic review. *J Alzheimers Dis*. 2007;12(1):23-35.
16. Grundman M, Petersen RC, Ferris SH, et al. Mild Cognitive Impairment can be

- distinguished from Alzheimer Disease and Normal Aging for clinical trials. *Arch Neurol*. 2004;61(1):59-66.
17. Manly JJ, Tang MX, Schupf N, Stern Y, Vonsattel JPG, Mayeux R. Frequency and course of Mild Cognitive Impairment in a multiethnic community. *Ann Neurol*. 2008;63(4):494-506. doi:10.1002/ana.21326.
 18. Apostolova LG, Cummings JL. Neuropsychiatric manifestations in Mild Cognitive Impairment: a systematic review of the literature. *Dement Geriatr Cogn Disord*. 2008;25(2):115-126. doi:10.1159/000112509.
 19. Plassman BL, Williams JW, Burke JR, Holsinger T, Benjamin S. Factors associated with risk for and possible prevention of cognitive decline in later life. *Ann Intern Med*. 2010;153:182-193.
 20. Tricco AC, Soobiah C, Berliner S, et al. Efficacy and safety of cognitive enhancers for patients with Mild Cognitive Impairment: a systematic review and meta-analysis. *CMAJ*. 2013;185(16):1393-1401. doi:10.1503/cmaj.130451.
 21. Reijnders J, van Heugten C, van Boxtel M. Cognitive interventions in healthy older adults and people with Mild Cognitive Impairment: a systematic review. *Ageing Res Rev*. 2013;12(1):263-275. doi:10.1016/j.arr.2012.07.003.
 22. Association A. Stages of Alzheimer's. http://www.alz.org/alzheimers_disease_stages_of_alzheimers.asp. Published 2017. Accessed August 23, 2017.
 23. Skinner J, Carvalho JO, Potter GG, Thames A, Zelinski E. The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): an expansion of the ADAS-Cog to improve responsiveness in MCI. *Brain Imaging Behav*. 2012;6:489-501. doi:10.1007/s11682-012-9166-3.
 24. Raghavan N, Samtani MN, Farnum M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimer's Dement*. 2013;9(10):S21-S31. doi:10.1016/j.jalz.2012.05.2187.
 25. Lundin-Olsson L, Nyberg L, Gustafson Y, Himbert D, Seknadj P, Karila-cohen D. "Stops walking when talking" as a predictor of falls in elderly people. *Lancet*. 1997;349:617.
 26. Buracchio T, Dodge HH, Howieson D, Wasserman D, Kaye J. The trajectory of gait speed preceding Mild Cognitive Impairment. *Arch Neurol*. 2010;67(8):980-986. doi:10.1001/archneurol.2010.159.
 27. Montero-Odasso M, Oteng-Amoako A, Speechley M, et al. The motor signature of Mild Cognitive Impairment: results from the Gait and Brain Study. *Journals Gerontol Ser A Biol Sci Med Sci*. 2014;69(11):1415-1421. doi:10.1093/gerona/glu155.
 28. Kueper JK, Speechley M, Lingum NR, Montero-Odasso M. Motor function and incident dementia: a systematic review and meta-analysis. *Age Ageing*. 2017:1-10. doi:10.1093/ageing/afx084.
 29. Montero-Odasso M. Gait as a biomarker of cognitive impairment and dementia syndromes. Quo vadis? *Eur J Neurol*. 2016;23(3):437-438. doi:10.1111/ene.12908.
 30. Montero-Odasso M, Verghese J, Beauchet O, Hausdorff JM. Gait and cognition: a complementary approach to understanding brain function and the risk of falling. *J Am Geriatr Soc*. 2012;60(11):2127-2136. doi:10.1111/j.1532-5415.2012.04209.x.
 31. Beauchet O, Allali G, Montero-Odasso M, Sejdíć E, Fantino B, Annweiler C. Motor phenotype of decline in cognitive performance among community-dwellers

- without dementia: population-based study and meta-analysis. *PLoS One*. 2014;9(6). doi:10.1371/journal.pone.0099318.
32. Verghese J, Annweiler C, Ayers E, et al. Motoric Cognitive Risk Syndrome multicountry prevalence and dementia risk. *Neurology*. 2014;83(8):718-726. doi:10.1212/WNL.0000000000000717.
 33. Montero-Odasso MM, Barnes B, Speechley M, et al. Disentangling Cognitive-Frailty: results from the Gait and Brain Study. *J Gerontol A Biol Sci Med Sci*. 2016;71(11):1476-1482. doi:10.1093/gerona/glw044.
 34. Montero-Odasso M, Hachinski V. Preludes to brain failure: executive dysfunction and gait disturbances. *Neurol Sci*. 2014;35(4):601-604. doi:10.1007/s10072-013-1613-4.
 35. Allali G, Assal F, Kressig RW, Dubost V, Herrmann FR, Beauchet O. Impact of impaired executive function on gait stability. *Dement Geriatr Cogn Disord*. 2008;26(4):364-369. doi:10.1159/000162358.
 36. Allali G, Annweiler C, Blumen H, et al. Gait phenotype from MCI to moderate dementia: results from the GOOD initiative. *Eur J Neurol*. 2016;23(3):527-541. doi:10.1111/ene.12882.
 37. Annweiler C, Beauchet O, Bartha R, et al. Motor cortex and gait in Mild Cognitive Impairment: a magnetic resonance spectroscopy and volumetric imaging study. *Brain*. 2013;136(3):859-871. doi:10.1093/brain/aws373.
 38. Welmer AK, Rizzuto D, Qiu C, Caracciolo B, Laukka EJ. Walking speed, processing speed, and dementia: a population-based longitudinal study. *J Gerontol A Biol Sci Med Sci*. 2014;69(12):1503-1510. doi:10.1093/gerona/glu047.
 39. Hausdorff JM, Buchman AS. What links gait speed and MCI with dementia? A fresh look at the association between motor and cognitive function. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2013;68(4):409-411. doi:10.1093/gerona/glt002.
 40. Muir SW, Speechley M, Wells J, Borrie M, Gopaul K, Montero-Odasso M. Gait assessment in Mild Cognitive Impairment and Alzheimer's Disease: the effect of dual-task challenges across the cognitive spectrum. *Gait Posture*. 2012;35(1):96-100. doi:10.1016/j.gaitpost.2011.08.014.
 41. Montero-Odasso MM, Sarquis-Adamson Y, Speechley M, et al. Association of dual-task gait with incident dementia in Mild Cognitive Impairment. *JAMA Neurol*. 2017;54(3):S154. doi:10.1001/jamaneurol.2017.0643.
 42. Hausdorff JM, Yogev G, Springer S, Simon ES, Giladi N. Walking is more like catching than tapping: gait in the elderly as a complex cognitive task. *Exp Brain Res*. 2005;164(4):541-548. doi:10.1007/s00221-005-2280-3.
 43. Lusardi MM, Pellecchia GL, Schulman M. Functional performance in community living older adults. *J Geriatr Phys Ther*. 2003;26(3):14-22.
 44. Adams K, Bayliss E, Blumenthal D, et al. Universal health outcome measures for older persons with multiple chronic conditions. *J Am Geriatr Soc*. 2012;60(12):2333-2341. doi:10.1111/j.1532-5415.2012.04240.x.

Chapter 2

2 Introduction to Outcome Measures

The purpose of Chapter 2 is to review important concepts and terminology related to the development and use of outcome measures for health conditions involving latent traits, to outline approaches for improving pre-existing outcome measures, and to introduce three of the main cognitive outcome measures used today.

The overall goal of a health-related outcome measure is to score specific traits to help determine whether a health condition is present, or to assess the relative severity of that health condition in an individual or group. The approaches and challenges differ depending on the nature of the outcome, specifically whether it is a manifest variable (e.g. a physical property, such as gait speed), or a latent trait (e.g. cognitive ability). Because manifest variables are often measured directly with instruments and devices, evaluating measurement is primarily a technical exercise concerned with reliability, accuracy, and precision. Latent traits are more difficult to evaluate. Section 2.1 will provide a brief overview of what latent traits are and how they can be modelled, and then Sections 2.2 and 2.3 will describe two main measurement models used to assess measures of latent traits, namely Classical Test Theory (CTT) and Item Response Theory (IRT).

2.1 Overview of Latent Traits

In contrast to medical abnormalities that can be physically seen or detected, such as a broken bone, there are many health states or conditions, often with a strong psychological component, which cannot be directly observed.^{1,2} Rather, they are associated with some underlying ability that is not directly observable, such as cognition or personality, that drives people to behave or function in certain ways.^{1,2} Outcome measures can quantify these latent traits using multiple test items that capture observable manifestations of the latent traits.³⁻⁵ The covariation between a subject's observed test item responses is assumed to be due to the latent trait.¹⁻⁴ Latent traits can be modelled in three main ways:

- 1) **Categorical** latent traits include discrete, mutually exclusive classes, that can be

used to separate a group of people who may appear similar based on observable traits (Figure 1).^{4,6} Members within each class share the same latent trait category, and are considered homogenous.^{4,6}

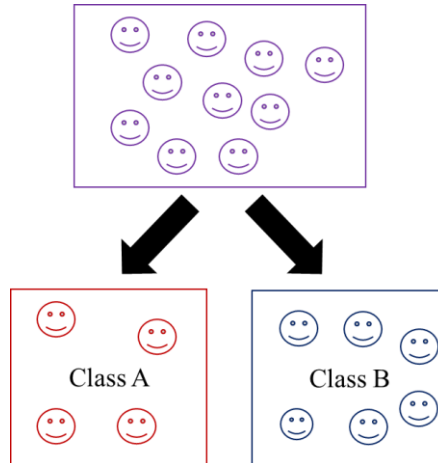


Figure 1 Categorical latent trait.

- 2) **Dimensional** latent traits follow a single continuum spanning from low to high magnitude of the latent trait (Figure 2).^{4,6} Subjects can be given a quantitative score to indicate their placement on the continuum, and then compared to each other (e.g. Subject A has poorer short-term memory than Subject B), but no straightforward group classification is available.⁶

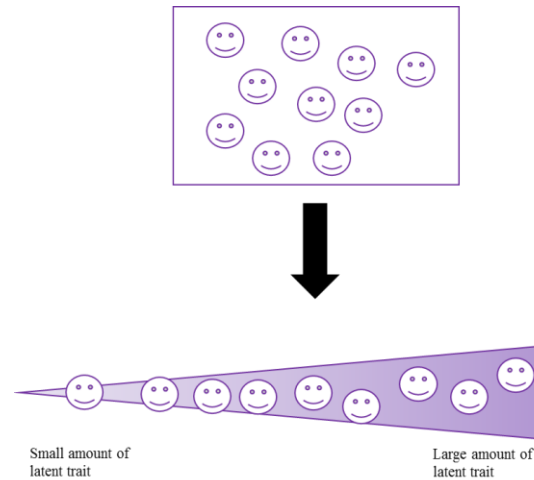


Figure 2 Dimensional latent trait.

- 3) **Factor Mixed Model** latent traits include both a categorical and a dimensional structure (Figure 3).⁶ Factor mixed model latent traits categorize subjects into different latent classes (e.g. subject has Subtype A of Disease X), and within each latent class subjects can be organized along a latent trait continuum (e.g. to indicate within-class differences in level of disease severity).⁶ Characteristics of the dimensional latent traits may differ between categorical latent classes.⁶

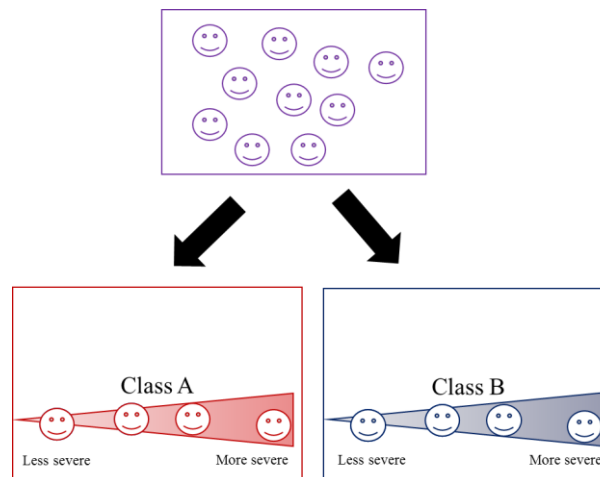


Figure 3 Factor mixed model latent trait.

Unidimensionality refers to the situation where one dimensional latent trait is responsible for all scores produced by an outcome measure or test (Figure 4).^{2-4,7} More specifically, the probability of responding a certain way on a test item is assumed to be a function of the underlying trait.^{4,6} The shape of the probability function will depend both on the underlying trait and on the format of the test question.⁵ For a unidimensional outcome measure where all items are designed to measure the same underlying trait, scores can be used to compare subjects' relative abilities on the underlying trait.⁸

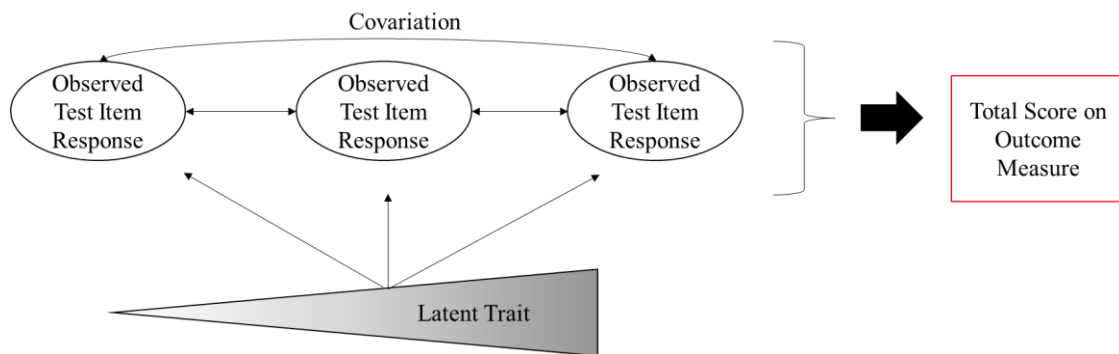


Figure 4 Unidimensionality assumption.

2.2 Classical Test Theory

CTT is one of two main psychometric theories underlying outcome measurement, and is focused on the observed scores of an outcome measure.⁹ A subject's observed score on any single measure administration is assumed to be composed of their "true score" and some error of measurement.^{2,8,9,10} True scores are sometimes referred to as "trait scores" as they are intended to relate to a subject's latent trait ability.⁸ CTT maintains an assumption of unidimensionality.^{2,10} Measurement errors are assumed to be random, follow a normal distribution with mean of zero, and not correlate with the true score.^{2,5,9,10}

$$X_i = T_i + E_i$$

Where X_i = Subject i 's observed total score on an outcome measure, T_i = Subject i 's true score or latent ability level, E_i = error of measurement.

Observed total scores are often obtained using an unweighted sum of all individual item responses on an outcome measure.² This method assumes that all items are equally difficult and equally important, that the difference between any two response options is the same across all test items, that all items correlate equally with the latent trait, and that subjects have responded to all items on an outcome measure.^{2,10} Rarely are all assumptions of CTT met, but the CTT model cannot be disproved because the assumptions cannot be directly tested against an unknown latent trait.⁹

Another limitation of CTT is the assumption of measurement invariance.² An outcome measure is said to be invariant when it performs the same way regardless of what, or who, is being measured, because the method by which results are produced is independent of the individual object, construct, or person being assessed in any given testing situation.^{11,12} For example, a scale designed to assess Attribute X should be able to identify the same amount of Attribute X in two people who truly do have the same amount of Attribute X, but differ by age or education. In reality, properties of outcome measures constructed using CTT are dependent on the samples in which they were tested and validated.¹⁰ Thus, with CTT there is a circular dependency between outcome measure properties and subject attributes.¹⁰ While properties of the outcome measure such as reliability and validity depend on the sample composition, especially how homogenous the sample is, subject scores depend on the properties of the outcome measure.^{2,10} A final limitation of CTT is that different outcome measure scores obtained under CTT cannot be compared unless transformed to Z-scores, T-scores, or percentiles.¹⁰ This method of transformation requires the raw outcome measure scores to be approximately normally distributed.¹³

2.2.1 Standardization

Any random variable following a normal distribution, which is a bell-shaped probability distribution, can be standardized by subtracting the mean and dividing by the standard deviation so that it becomes a standard normal distribution with a mean of zero and standard deviation of one.¹³ The resulting standard scores, or Z-scores, can be expressed as percentiles and allow direct comparison of scores from outcome measures which

initially had different scoring methods, scales, means, and standard deviations.¹⁰

2.2.2 Reliability

Reliability is the ability of an outcome measure to give the same results for the same subjects, under different circumstances, assuming the underlying construct of interest has remained constant.^{10,14} Four main subtypes of reliability include:

- 1) **Inter-observer reliability** indicates the degree to which two different people administering the same outcome measure to the same subject at the same time, will produce the same results.^{2,10,14}
- 2) **Intra-observer reliability** indicates the degree to which the same person administering the same outcome measure in two different circumstances, will produce the same results.¹⁰
- 3) **Test-retest reliability** refers to the ability of an outcome measure to give the same results for the same subject at two different time points, assuming the subject remained stable for whatever the measure was designed to assess.^{2,10,14} If a subject has changed with regards to the construct of interest, then an outcome measure designed to assess change with high reliability should reflect this change in the final score.^{2,15}
- 4) The first three types of reliability apply to both single items and to scores based on several items. **Internal consistency** is applicable only to composite scores, and refers to whether all items of an outcome measure assess the same construct.^{2,10,14}

There are several approaches that can be used to assess reliability, and an outcome measure which shows high reliability for a certain population and context of assessment may not demonstrate similarly high reliability for a different population or context.^{2,10,15} Hence, it is important to refer to reliability of test scores in specified populations and contexts, not the reliability of an outcome measure on its own.^{2,10} Reliability parameters range from zero to one, and will increase as between-subject variation increases and decrease as measurement error decreases^{2,10,16}:

$$\text{Reliability} = \sigma^2_{\text{subject}} / [\sigma^2_{\text{subject}} + \sigma^2_{\text{error}}]$$

Where $\sigma^2_{\text{subject}}$ = True variance between subjects, σ^2_{error} = Measurement error variance.

2.2.3 Precision

Precision in the context of outcome measurement refers to the reproducibility of a score for a given subject in a given circumstance.¹⁷ Although similar to reliability, precision does not distinguish between score variability due to true subject differences and variability due to error.¹⁷ An outcome measure may demonstrate high precision but low reliability.^{10,17} For example, if a group of subjects all obtain very similar scores on an outcome measure there will be very little true subject variability and therefore low reliability, but high precision.¹⁷

2.2.4 Standard Error of Measurement

The Standard Error of Measurement (SEM) provides an absolute measure of the precision of individual subject scores, expressed in the same units as the outcome measure.^{10,16} As described above, under CTT the score on any single outcome measure administration consists of the true test score and measurement error.^{2,8-10} If an infinite number of test administrations were performed for a subject, the average of the observed test scores would be the best single estimate of the true score for that subject.^{5,10,18,19} The true score in this sense is referring to consistency, not validity.¹⁹ The standard deviation of the sampling errors for the distribution of observed test scores from the hypothetical infinite number of administrations is the SEM.^{10,19} The SEM can be calculated as^{10,16}:

$$\text{SEM} = \sigma_x \sqrt{1 - R} = \sqrt{\sigma^2_{\text{error}}}$$

Where σ_x = Standard deviation of the observed scores over a population, R = Reliability, σ^2_{error} = Measurement error variance.

2.2.5 Coefficient of Variation

The coefficient of variation (CV) is a measure of intra-individual variability that accounts

for the overall performance of study participants on an outcome measure, calculated as^{20,21}:

$$CV = \sigma / \mu$$

Where σ = standard deviation, μ = mean.

The CV is dimensionless and allows comparison of score variability for the same outcome measure administered to different samples, or comparison of variability for different outcome measures administered to the same or different samples.²¹

2.2.6 Validity

In general, validity is the extent to which an outcome measure evaluates what it was designed to measure, and is specific to both the population and context of assessment.^{2,10,18} Validity can be expressed as¹⁰:

$$\text{Validity} = \sigma^2_{\text{construct of interest}} / \sigma^2_{\text{observed}}$$

Where $\sigma^2_{\text{observed}} = \sigma^2_{\text{construct of interest}} + \sigma^2_{\text{systematic error}} + \sigma^2_{\text{random error}}$

Validity exists on a continuum, and whether or not a measurement tool is “valid” for a particular population and context requires a decision based on results from a series of hypothesis tests that make up the process of validation.^{10,18} As knowledge of a health condition or state increases, or the theoretical framework underpinning a construct of interest changes, further validation will be required to determine whether a previously developed outcome measure remains valid enough for use in the current context and population of interest.¹⁰ If an outcome measure is designed to assess a multidimensional construct, then validation for each of the individual dimensions must be performed using separate hypotheses pertaining to each dimension.^{2,19} Three major subdomains of the validation process include:

- 1) **Content validation** determines the extent to which an outcome measure assesses all important components of a health condition of interest.^{10,15,18,22} As more components are captured by an outcome measure, inferences about the true

underlying health of a subject obtaining a given score can become more comprehensive.^{10,15,22}

Content validation differs from other types of validity in two important ways. First, an outcome measure can have high content validity even with low reliability.¹⁰ Content validity is improved by including items to assess all important aspects of a health condition, but if these aspects are highly variable across individuals internal consistency may be compromised.¹⁰ Secondly, content validation is a non-empirical approach as it depends solely on the judgement of experts in the field or comparison with theoretical models, rather than on statistical tests of comparison with other measures of the health condition of interest.^{10,15,18,22}

- 2) **Criterion validation** assesses how well an outcome measure agrees with other well-established measures of the same health condition.^{10,15,18,22} Subtypes of criterion validation include concurrent validation, whereby the comparison between the two measures is made at the same point in time, and predictive validation, where the outcome measure under study is compared to some measured criteria that occurs in the future.^{18,22}
- 3) **Construct validation** assesses whether an outcome measure outperforms (rather than mimics as with criterion validation) the gold standard, or criteria currently believed to be the best possible assessment of a health condition of interest.^{10,15,22,23} Unlike criterion and content validation, which can be estimated for a particular population and context in a single appropriate study, construct validation is a continuous process and tests the theory and outcome measure at the same time.¹⁰ Construct validation requires both justification with explicit reference to evidence about the relevance of the components of an outcome measure to the health condition of interest, as well as statistical tests and numerical comparisons with measures of the health condition or individual components of it.^{22,23} As the theoretical framework of a health condition changes, hypotheses may change and further tests need to be conducted.^{10,22,23}

Construct validation may be further broken down into convergent and discriminant validation. *Convergent validation* tests how closely an outcome measure relates to other variables and measures of the construct it was designed to measure.¹⁰ *Discriminant* or *divergent validation* tests that the outcome measure undergoing validation does not correlate with measures of constructs hypothesized to not be part of the health condition of interest.¹⁰ An acceptable strength of correlation between the outcome measure undergoing validation and the other pre-existing measures will depend on the relative importance of what is being assessed by the other measures for the health condition of interest.¹⁰

2.2.7 Responsiveness

Responsiveness is a type of validity, however for simplicity this thesis will review responsiveness as a separate concept.^{10,19} Responsiveness is broadly defined as the ability of an outcome measure to accurately detect change,^{15,24-26} and must be contextualized by the type of change being assessed.^{24,26} This contextualization may occur according to the taxonomy of responsiveness developed by Beaton et al. (2001), which includes three axes of classification:

- 1) The **“Who” axis** differentiates between individual level and group level of analysis and interpretation.²⁶
- 2) The **“Which” axis** describes whether the scores being contrasted are measuring between-person differences at one point in time, within-person changes over time, or between-person differences of within-person change over time.²⁶
- 3) The **“What” axis** specifies the type of change being quantified in the study, such as minimum potentially detectable change by the instrument, observed change measured by an instrument in a population, or observed change in a population deemed to have improved by a clinician.²⁶

The three conceptualizations of change most relevant for this thesis, and examples of methods for assessing responsiveness for each, are presented below. Please note that the group-level analysis and interpretation of change is often used for research studies, but

outcome measures will require adequate levels of responsiveness to individual-level change if they are intended to also be used for one-on-one assessments, such as in a clinical setting.

Baseline Discrimination: *Responsiveness to group-level between-person differences in stage of disease progression at one point in time.* The health condition of interest for this thesis can be viewed as a continuum of severity which includes the key stages of Normal Cognition (NC), Subjective Cognitive Impairment (SCI), Mild Cognitive Impairment (MCI), and dementia.²⁷⁻³³ At any arbitrary baseline point, subjects at the different stages of disease progression beyond NC have already changed in terms of their underlying disease pathology and phenotypic expressions of such. And, if the natural history of the disease has predictable stages, then all subjects are expected to go through similar changes in phenotypic expression as they progress from NC to severe dementia. Just because the changes did not occur within the observation window of the study does not mean it is not change we are measuring – it means the change has occurred retrospectively. Therefore, the ability of a measurement tool to discriminate between subjects with NC, SCI, MCI, dementia, or even more refined categories at one point in time, can be interpreted as a type of responsiveness. This separation of subjects into distinct diagnostic categories is adopting a categorical latent trait conceptualization of the dementing process.

A simple way to assess baseline discrimination is to compare the mean scores of the outcome measure in each of the predefined diagnostic categories. These scores may be tested for statistically significant differences with a *t-test* or *Analysis of Variance*, depending on the number of groups to be compared.¹³ Non-parametric counterparts to these tests, *Mann Whitney U* and *Kruskal-Wallis*, are also suitable options.¹³

It is important to note that if the health condition of interest did not progress continuously through stages of severity in its natural history, it would not be appropriate to refer to baseline discrimination as a type of responsiveness. For example, if we were trying to identify children with different allergies, a child with a peanut allergy is not expected to have previously been in a citrus allergy category, nor are they expected to progress towards a fish allergy category. For these types of health conditions defined with nominal

categories, baseline discrimination is more similar to sensitivity, with the caveat that discriminative ability is often quantified in relation to other measures while sensitivity is compared to pre-specified criteria of presence versus absence.¹⁰ Sensitivity is further described in Section 2.2.8.

Disease progression: *Responsiveness to group-level within-person observed change measured by an outcome measure in a given population.* This may include progression (i.e. increasing severity) within or between the above-mentioned stages of dementia progression.²⁶ It is not limited by those diagnostic categories, and therefore is adopting a dimensional latent trait conceptualization of the entire dementing process.

The main statistical tests used to assess this type of responsiveness include *paired t-tests*, which test the null hypothesis that there is no change in the individual outcome scores for the same group across two time points, *standardized effect sizes*, which express the magnitude of change in outcome measure scores across two time points by comparing the average amount of within-person change to the variability of baseline scores, and the *standardized response mean (SRM; also known as the signal-to-noise ratio, responsiveness-treatment coefficient, efficiency index, or standardized change)*, which expresses the magnitude of change in the observed outcome measure scores relative to the variability of those change scores.²⁴ A key advantage of the SRM is that it allows direct comparison between different outcome measures because it takes into account the fact that different measures have different score ranges and variability in change from baseline.³⁴

Treatment Effect: *Responsiveness to group-level between-person differences of within-person observed change over time.* Sample statistical tests that may be used to assess this type of responsiveness are a *two-way Analysis of Variance* including a treatment group by time factor, an *Analysis of Covariance* with terms for baseline score and treatment group, or *regression* models.¹³ The scores from an outcome measure designed to be responsive to a treatment effect can be used to calculate an *effect size*, which summarizes the magnitude and direction of differences between two or more groups which differ on at least one important characteristic (in this case, whether or not the group received active treatment).³⁵ For example, a treatment effect in a clinical trial may be that the group

receiving active treatment improved their score on the outcome measure by 15% whereas the group receiving placebo remained stable. Effect sizes can also include different rates of deterioration, such as if the treatment group experienced average 5% worsening in scores compared to the placebo group's average 15% worsening. Regardless of statistical significance, the clinical significance of a treatment effect size should always be interpreted within the context of what is being assessed and how much of a reduction (or increase) in the outcome is meaningful to patients, caregivers, or clinicians.³⁵ Effect sizes can be adjusted for potential confounding factors. Confounders may increase or decrease the magnitude of effect, or change the direction of effect (qualitative confounding).³⁵ This is important to keep in mind when comparing effect sizes from different studies.

2.2.8 Sensitivity and Specificity

Sensitivity and specificity are characteristics of outcome measures designed to classify a subject as having or not having a health condition.³⁶⁻³⁸ These outcome measures do not need to be based on CTT. Sensitivity and specificity are calculated from the perspective of a 'gold standard' or external criterion:

Sensitivity, or true positive probability, is the ability to detect a specific health condition, when that health condition is truly present.^{10,36-38}

$$\text{Sensitivity} = \# \text{ True Positives} / (\# \text{ True Positives} + \# \text{ False Negatives})$$

Specificity, or true negative probability, is the ability to identify those without a specific health condition, when that health condition is truly absent.^{10,36-38}

$$\text{Specificity} = \# \text{ True Negatives} / (\# \text{ True Negatives} + \# \text{ False Positives})$$

Where True Positive = The test provides a positive result (disease present) when the subject really does have the health condition of interest, False Positive = The test provides a positive result when the subject does not really have the health condition of interest, True Negative = The test provides a negative result (disease absent) when the subject really does not have the health condition of interest, False Negative = The test provides a negative result when the subject really does have the health condition of interest. Overall

accuracy can be obtained by dividing the number of correct assessments (True Positives + True Negatives) by the total number of assessments.³⁸

When the health condition of interest exists on a continuum of severity that can be captured by increasing or decreasing scores on an outcome measure, cut-points can be used to decide what score corresponds to a positive test result indicating that a subject may have the health condition of interest.¹⁰ Choosing higher or lower cut-points will alter the sensitivity and specificity of the test. Usually when sensitivity increases, specificity decreases, and vice versa.^{10,37} The prevalence of the health condition in the population being tested does not affect the sensitivity and specificity.^{10,37}

Receiver Operating Characteristic (ROC) curves plot false positives (1-specificity) against true positives (sensitivity) for all possible cut-off values.^{38,39} The area under the curve (AUC) represents the probability that, given a random pair of people where one truly has the outcome and one does not, the person who has the outcome will score higher than the person who does not.^{10,39} The AUC ranges from 0.5 (random chance that the test will correctly classify a patient) to 1.0 (test perfectly classifies all patients).^{37,38} ROC curves can be used to assess responsiveness when responsiveness is described in terms of sensitivity and specificity for detecting change and no change in an external standard. The external standard score must be dichotomized at a cut-off for what constitutes meaningful change.

2.3 Item Response Theory

The above definitions of reliability, validity, and responsiveness are based on CTT.¹⁰ The second major measurement model used to assess outcome measures for latent traits is IRT.^{8,10} IRT is both a measurement model and a probability model.^{8,9} It estimates the probability of a subject selecting a particular test item response given their ability on a latent trait.^{8,10} So, unlike CTT which focuses on the total test score, IRT focuses on individual test items.^{9,10} Furthermore, IRT does not assume that all test items are equivalent.^{8,10} Rather, specific item properties can be built into the IRT model to try to obtain the best possible estimate of a subject's level of latent trait ability.^{8,10}

IRT models are usually described by the number of item parameters they contain. Three different item parameters have been defined, any number of which may be included in an IRT model.^{8,10} A multi-item outcome measure may contain items which have the same or different values for each of the item parameters.⁸ The three item parameters can be described visually with item characteristic curves (ICC) or item characteristic functions in plots of the latent trait ability against the probability of a particular response (Figure 5).^{8,10} ICCs can be plotted for dichotomous or polytomous items, and each item response option may be given its own ICC on the plot.⁸

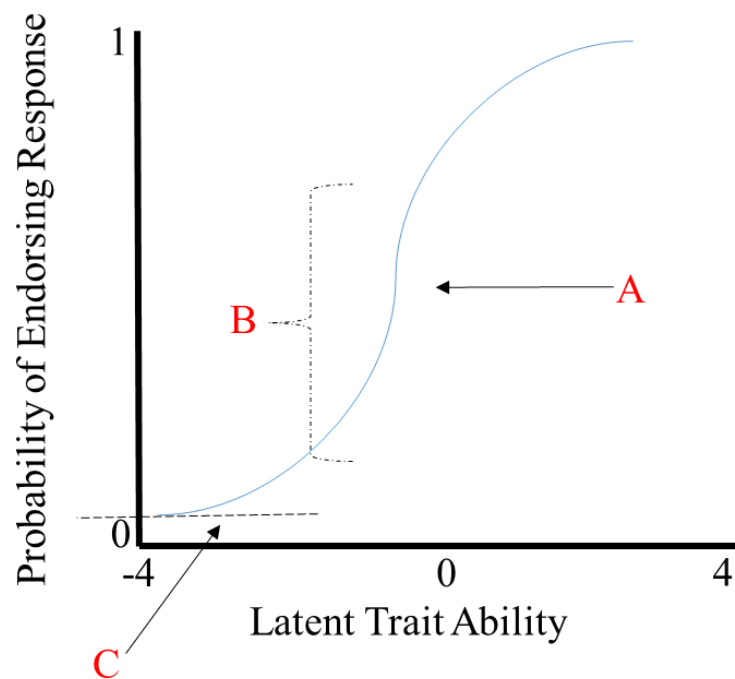


Figure 5 Sample item characteristic curve.

The first item parameter is an **item difficulty parameter**, and is situated at the point of inflection of the ICC, or the point at which the probability of selecting a particular response option is 0.5 (Figure 5, point A).^{8,10} The purpose of the item difficulty parameter is to locate each test item on the same continuum of latent trait ability that subjects are located on.^{8,10}

The second is an **item discrimination parameter**, and is reflected by the slope of the

ICC (Figure 5, point B).^{5,8,10} The item discrimination parameter relates to how much information a test item holds about the underlying latent ability.^{8,10} An item with a steeper slope is better able to discriminate between levels of latent ability, for the levels of ability which it covers, because the probability of the response changes very quickly as one moves along the latent trait continuum.^{5,8,10}

The third is an **item guessing parameter**, identified by the lower asymptote of the ICC (Figure 5, point C).⁸ The item guessing parameter models the chance probability of responding to a test item in a certain way (e.g. how likely one is to choose a correct response by guessing).^{8,10}

Model-data fit analyses can be used to test whether an IRT model is a good description of the data, whereas with CTT the model is just assumed to be true.⁹ IRT also does not calculate reliability in the same way as CTT.¹⁰ IRT focuses on precision rather than reliability, whereby higher precision indicates a higher level of “information”.^{8,10}

$$\text{Standard Error (latent trait)} = 1 / \sqrt{\text{information (latent trait)}}$$

Methods for calculating level of information differ among IRT models, but in general higher information corresponds to lower standard error of estimate for a person’s location on the latent trait continuum, and the higher the item discrimination parameter, the higher the item information.^{8,10} Item information is the cumulative sum of all information from all of that item’s response categories.⁸ Test information is the cumulative sum of the information from all test items.⁸ The information peak is located at the apex of an ICC, and corresponds to the level of latent trait ability for which the item (or response option, or test) holds the most information.⁸

The simplest IRT model is the Rasch model, which is a particular type of one parameter logistic model.^{8,10} A one parameter logistic model provides the probability of a particular item response given the subject’s latent ability, the item difficulty, and the item discrimination.^{8,10} For a Rasch model the item discrimination parameter is set to 1, and for the one parameter logistic model it can be any number, but that number remains constant for all items.^{8,10} The two parameter logistic model builds on the one parameter

logistic model by allowing the parameter for test item discrimination to differ between items.^{8,10} The three parameter logistic model adds in the item guessing parameter.^{8,10} IRT models have three main assumptions:

- 1) **The dimensionality assumption** states that test responses are due to one's latent trait ability level.^{8,10} Although most often the latent trait is assumed to be unidimensional, modifications can be made for multidimensional cases. If different latent traits underlie different test items (between-item multidimensionality) then separate IRT models can be built for clusters of items relying on the same latent trait.⁸ If multiple latent traits underlie responses to a single item, then a multidimensional IRT model can be built.⁸ A multidimensional IRT model can be either compensatory or non-compensatory depending on whether one latent trait is able to compensate for deficiencies in the other latent trait.^{8,10}
- 2) **The conditional or local independence assumption** states that each individual item response is independent of any other item responses given the subject's latent trait ability.⁸⁻¹⁰ In other words, test item responses are due completely to a specific latent trait ability, not due to other latent traits or knowledge or priming from other test items.⁸⁻¹⁰
- 3) **The functional form assumption** includes whether the model correctly specifies the function that the data follow.^{8,10}

2.4 Approaches for Modifying Outcome Measures

Two main approaches for modifying, or attempting to improve, a pre-existing outcome measure are changing the scoring methodology and adding additional test items.⁴⁰ These approaches can be used individually or in combination.⁴⁰ Regardless of the approach taken, it is recommended that the modified measure be backwards-compatible with the original measure. Backwards compatibility means the original measure can be recovered from the new or modified version, which preserves the ability to compare results between studies using the modified and original version.⁴¹ When the measure to be modified is

considered the current gold standard, backwards compatibility will also allow researchers to show regulators results from their study in terms of the original gold standard measure, while results from the modified version can be used for further research purposes including demonstration of improved measurement properties.

2.4.1 Statistical Modification of Scoring Methodology

The simplest scoring method for any outcome measure is a straight summation of points across all test items, whereby points gained from any test item contribute equally to the total score.^{2,42} This transparent method assumes that all test items are equally important and informative for assessing disease severity.⁴³⁻⁴⁵ Item weights can be altered so that a point on one test item would contribute more or less to the total score than a point on a different test item.² Applying a weight of zero to a test item is effectively the same as removing the item from the scoring process, while maintaining the capacity for backwards compatibility. Values for the re-weighting process may be derived statistically or theoretically in accordance with the relative importance of each item for the health condition of interest. Psychometric methods, such as IRT, can also be used to modify scoring for any given outcome measure.^{43,45}

2.4.2 Adding Additional Item Content

The second approach to improve a pre-existing outcome measure is to administer the original outcome measure along with additional test items.^{34,40,46} Often items are added because theoretical advancements in a field identified an important component of a health condition that is not effectively being assessed by the original outcome measure. Statistical approaches to adding test items may also be used. The total score range of the modified outcome measure can be derived by simply extending the scoring range of the original version to accommodate the additional items, or by some other scoring modification. The additional item content may change the latent trait(s) assessed by the outcome measure, warranting the need for further validation studies of the additional item content as well as the modified outcome measure as a whole.¹⁰

2.5 Key Cognitive Outcome Measures

Three main cognitive outcome measures used today include the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MoCA), and the Alzheimer Disease Assessment Scale – Cognitive Subscale (ADAS-Cog).

The MMSE was developed in 1975 to address the need for an outcome measure to screen for possible cognitive impairment in several psychiatric conditions including but not limited to dementia.⁴⁷ The MMSE contains 11 questions, can be administered in five to ten minutes, and is scored from 0 to 30 with higher scores reflecting better cognitive function.⁴⁷ Domains of cognitive function assessed include orientation, memory, language, attention, and visuospatial abilities.^{47,48} Although not developed specifically for dementia, the MMSE is the most commonly used screening test for dementia, whereby a score of 23 or 24 is often selected as the cut-off to identify subjects with probable dementia.⁴⁸ A meta-analysis of the diagnostic performance of the MMSE across 108 cohort studies with different patient populations found that the overall summary sensitivity of the MMSE for detecting dementia was 0.81 (95% Confidence Interval (CI) 0.78, 0.84), the specificity 0.89 (95% CI 0.87, 0.91), and the overall accuracy 92% (95% CI 90, 94).⁴⁸ Twenty-one cohorts in the same meta-analysis assessed the ability of the MMSE to detect MCI, yielding a summary sensitivity of 0.62 (95% CI 0.52, 0.71), and specificity of 0.87 (95% CI 0.80, 0.92).⁴⁸ This comparatively low performance for detecting MCI may be because many people meeting the clinical criteria for MCI score in the “normal” range on the MMSE (over 26 points).⁴⁹

The MoCA was developed in 2005 for the purpose of improving detection of MCI specifically.⁴⁹ The MoCA contains 10 questions, can be administered in under 10 minutes, and is scored from 0 to 30 with higher scores indicating better cognitive function.^{48,49} Cognitive domains assessed by the MoCA include memory, executive function, attention, language, and orientation.^{48,49} The meta-analysis described above assessed the diagnostic performance of the MoCA for detecting MCI across nine cohorts, and found a summary sensitivity of 0.89 (95% CI 0.84, 0.92) and specificity of 0.75 (95% CI 0.62, 0.85).⁴⁸

While the MMSE and the MoCA were developed as screening tools to assist primary care physicians identify patients with cognitive difficulties, they are also used for group-level analyses in research studies. The ADAS-Cog is another commonly used cognitive outcome measure for studies of dementia and pre-dementia syndromes. In contrast to the MMSE and MoCA which were developed to identify people from a heterogeneous sample whom may have MCI or dementia, the ADAS-Cog was developed for the purpose of identifying severity of dysfunction in samples of subjects with known AD.⁵⁰ The ADAS-Cog is the current ‘gold standard’ for clinical trials of treatments for dementia, and is often also used as such in studies of MCI and other pre-dementia syndromes. Unfortunately, several concerns about the use of the ADAS-Cog have emerged. These concerns will be a main focus of this thesis. Most notably, the ADAS-Cog appears to have poor responsiveness to important changes in subjects with pre-dementia syndromes, and has low content validity since advancements in the study of dementia and pre-dementia syndromes have identified domains not covered by the ADAS-Cog which are emerging as important components of dementia and pre-dementia syndromes. Chapter three will provide a comprehensive literature review on the ADAS-Cog, including a description of individual scale tasks and scoring.

2.6 Summary

Latent traits are underlying dimensions, such as cognitive ability, which people possess but cannot be observed directly. Latent traits can be divided into categorical, dimensional, and factor mixed model structures. Cognition can be conceptualized using any of these structures, depending on the theoretical framework and measurement model being used. Two main measurement models used to develop outcome measures for latent traits include CTT and IRT. Three key cognitive outcome measures include the MMSE, MoCA, and ADAS-Cog. All three were developed using CTT. The ADAS-Cog is the current ‘gold standard’ outcome measure for clinical trials in dementia and pre-dementia syndromes, however there is some concern about its utility in studies of pre-dementia syndromes.

2.7 References

1. Cano SJ, Posner HB, Moline ML, et al. The ADAS-cog in Alzheimer's Disease clinical trials: psychometric evaluation of the sum and its parts. *J Neurol Neurosurg Psychiatry*. 2010;81:1363-1368. doi:10.1136/jnnp.2009.204008.
2. DeVellis RF. Classical Test Theory. *Med Care*. 2006;44(11):S50-S59.
3. Hambleton RK, Cook LL. Latent trait models and their use in the analysis of educational test data. *J Educ Meas*. 1977;14(2):75-96.
4. McCutcheon A. *Latent Class Analysis*. Newbury Park, California: Sage Pub; 1987.
5. Lord FM. The relation of test score to the trait underlying the test. *Educ Psychol Meas*. 1953;13:517-549. doi:10.1177/001316445301300401.
6. Clark SL, Muthén B, Rose RJ. Models and strategies for factor mixture analysis: an example concerning the structure underlying psychological disorders. 2013;20(4):1-29. doi:10.1080/10705511.2013.824786.Models.
7. Hattie JA. Methodology review: assessing unidimensionality of tests and items. *Appl Psychol Meas*. 1985;9(2):139-164. doi:10.1177/014662168500900204.
8. Ayala RJ. *The Theory and Practice of Item Response Theory*. New York: The Guilford Press; 2009.
9. Tractenberg RE. Classical and modern measurement theories, patient reports, and clinical outcomes. *Contemp Clin Trials*. 2010;31(1):1-2. doi:10.1016/S1551-7144(09)00212-2.
10. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales*. 5th ed. Oxford University Press; 2015.
11. van de Schoot R, Lugtig P, Hox J. A checklist for testing measurement invariance. *Eur J Dev Psychol*. 2012;9(4):486-492. doi:10.1080/17405629.2012.686740.
12. Borsboom D. When Does Measurement Invariance Matter? *Med Care*. 2006;44(11):S176-S181.
13. D'Agostino RB, Sullivan LM, Beiser AS. *Introductory Applied Biostatistics*. Brooks and Cole; 2006.
14. Mottola CA. Assessing and enhancing reliability. *Decubitus*. 1992;5(6):42-44.
15. Krishner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis*. 1985;38(1):27-36.
16. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the sem. *J Strength Cond Res*. 2005;19(1):231-240.
17. Streiner DL, Norman GR. "Precision" and "accuracy": two terms that are neither. *J Clin Epidemiol*. 2006;59(4):327-330. doi:10.1016/j.jclinepi.2005.09.005.
18. Fullerton JT. Evaluation of research studies. Part IV: validity and reliability-concepts and application. *J Nurse Midwifery*. 1993;38(2):121-125.
19. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine*. Cambridge University Press; 2011.
20. Tractenberg RE, Pietrzak RH. Intra-individual variability in Alzheimer's Disease and cognitive aging: definitions, context, and effect sizes. *PLoS One*. 2011;6(4):e16973. doi:10.1371/journal.pone.0016973.
21. Abdi H. Coefficient of variation. *Encyclopedia Res Des*. 2010;2. doi:10.1007/978-3-642-04898-2_177.
22. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52:281-302.

23. Cronbach LJ. The two disciplines of scientific psychology. *Am Psychol.* 1957;12:671-684.
24. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness. *J Clin Epidemiol.* 2000;53:459-468.
25. De Bruin AF, Diederiks JPM, De Witte LP, Stevens FCJ, Philipsen H. Assessing the responsiveness of a functional status measure: The Sickness Impact Profile versus the SIP68. *J Clin Epidemiol.* 1997;50(5):529-540. doi:10.1016/S0895-4356(97)00047-4.
26. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol.* 2001;54(12):1204-1217.
27. Reisberg B, Gauthier S. Current evidence for Subjective Cognitive Impairment (SCI) as the pre-Mild Cognitive Impairment (MCI) stage of subsequently manifest Alzheimer's Disease. *Int Psychogeriatrics.* 2008;20(1):1-16. doi:10.1017/S1041610207006412.
28. Reisberg B, Prichep L, Mosconi L, et al. The pre-Mild Cognitive Impairment, Subjective Cognitive Impairment stage of Alzheimer's Disease. *Alzheimer's Dement.* 2008;4:S98-S108. doi:10.1016/j.jalz.2007.11.017.
29. Petersen RC, Donohue MC, Gamst AC, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI) clinical characterization. 2010:201-209.
30. Samtani MN, Raghavan N, Shi Y, et al. Disease progression model in subjects with Mild Cognitive Impairment from the Alzheimer's Disease Neuroimaging Initiative: CSF biomarkers predict population subtypes. *Br J Clin Pharmacol.* 2012;75(1):146-161. doi:10.1111/j.1365-2125.2012.04308.x.
31. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Kokmen E, Tangalos EG. Aging, memory, and Mild Cognitive Impairment. *Int psychogeriatrics.* 1997;9(Supp 1):65-69.
32. Manly JJ, Tang MX, Schupf N, Stern Y, Vonsattel JPG, Mayeux R. Frequency and course of mild cognitive impairment in a multiethnic community. *Ann Neurol.* 2008;63(4):494-506. doi:10.1002/ana.21326.
33. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild Cognitive Impairment: a concept in evolution. *J Intern Med.* 2014;275(3):214-228. doi:10.1111/joim.12190.
34. Podhorna J, Krahnke T, Shear M, E Harrison J. Alzheimer's Disease Assessment Scale-Cognitive Subscale variants in Mild Cognitive Impairment and mild Alzheimer's Disease: change over time and the effect of enrichment strategies. *Alzheimers Res Ther.* 2016;8(1):8. doi:10.1186/s13195-016-0170-5.
35. Durlak JA. How to select, calculate, and interpret effect sizes. *J Pediatr Psychol.* 2009;34(9):917-928. doi:10.1093/jpepsy/jsp004.
36. Altman DG, Bland JM. Diagnostic tests 1 sensitivity and specificity. *BMJ.* 1994;308:1552-1553. doi:10.1258/phleb.2012.012J05.
37. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Contin Educ Anaesthesia, Crit Care Pain.* 2008;8(6):221-223. doi:10.1093/bjaceaccp/mkn041.
38. Šimundić A-M. Measures of diagnostic accuracy: basic definitions. *EJIFCC.* 2009;19(4):203-211.
39. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Casp J Intern Med.* 2013;4(2):627-635. doi:10.1017/CBO9781107415324.004.

40. Skinner J, Carvalho JO, Potter GG, Thames A, Zelinski E. The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): an expansion of the ADAS-Cog to improve responsiveness in MCI. *Brain Imaging Behav.* 2012;6:489-501. doi:10.1007/s11682-012-9166-3.
41. Clement C. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthop.* 2016;87(1):88-90. doi:10.3109/17453674.2015.1120127.
42. Hobart J, Cano S, Posner H, et al. Putting the Alzheimer's cognitive test to the test I: traditional psychometric methods. *Alzheimer's Dement.* 2013;9(1):S4-S9. doi:10.1016/j.jalz.2012.08.005.
43. Verma N, Beretvas SN, Pascual B, Masdeu JC, Markey MK. New scoring methodology improves the sensitivity of the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) in clinical trials. *Alzheimers Res Ther.* 2015;1-17. doi:10.1186/s13195-015-0151-0.
44. Wouters H, Gool WAVW van, Schmand B, Lindeboom R. Revising the ADAS-Cog for a more accurate assessment of cognitive impairment. *Alzheimer Dis.* 2008;0(0):1-9.
45. Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R. Putting the Alzheimer's cognitive test to the test II: Rasch Measurement Theory. *Alzheimer's Dement.* 2013;9(1):S10-S20. doi:10.1016/j.jalz.2012.08.006.
46. Raghavan N, Samtani MN, Farnum M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimer's Dement.* 2013;9(10):S21-S31. doi:10.1016/j.jalz.2012.05.2187.
47. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12(3):189-198. doi:0022-3956(75)90026-6.
48. Tsoi KKF, Chan JYC, Hirai HW, Wong SYS, Kwok TCY. Cognitive tests to detect Dementia: a systematic review and meta-analysis. *JAMA Intern Med.* 2015;175(9):1-9. doi:10.1001/jamainternmed.2015.2152.
49. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for Mild Cognitive Impairment. *JAGS.* 2005;53:695-699. doi:10.1111/j.1532-5415.2005.53221.x.
50. Rosen WG, Mohs RC, Davis KL. A New Rating Scale for Alzheimer's Disease. *Am J Psychiatry.* 1984;141(11):1356-1364. doi:10.1176/ajp.141.11.1356.

Chapter 3

3 Literature Review

The purpose of Chapter 3 is to explain how the Alzheimer's Disease (AD) Assessment Scale (ADAS) was developed, briefly review measurement properties of the ADAS Cognitive Subscale (ADAS-Cog) in dementia and pre-dementia populations, and provide a comprehensive review of all modifications that have been made to the ADAS-Cog and any assessments of the responsiveness of these modified versions. Please note that only literature published in the English language, and English language versions of the ADAS-Cog or modifications thereof, were examined.

3.1 Development of the Alzheimer's Disease Assessment Scale

The ADAS was originally designed to fulfill the need for a rating scale specific to AD studies. Goals for the ADAS included being able to assess the severity of cognitive and non-cognitive dysfunction from mild to severe dementia, while maintaining reliability and brevity of administration for subjects in different environments.¹

3.1.1 Item Selection

Item selection for the ADAS began with calculating the reliability and validity of forty candidate items in a development sample of 27 subjects with AD and 28 subjects with normal cognitive function (NC).¹ Most of the forty items showed statistically significant inter-rater reliability and test-retest reliability as assessed by intraclass correlation coefficients (ICC) and Spearman rank-order correlations, respectively, in the AD and NC groups separately.¹ Practice effects were detected only in the NC group.¹ Results from the AD group alone were used to select the final 21 items, which can be divided into cognitive and non-cognitive subscales. The ADAS takes about 45 minutes to administer. It is scored from 0 to 150 by summing the number of errors made on each test item so that

higher scores indicate worse performance.¹ The mean ADAS total score in the 27 subjects with AD was 37.0 (Standard Deviation (SD)=17.5).¹

The non-cognitive subscale (ADAS-Noncog) includes 10 assessments, scored from 0 to 50, which consider mood and behavioural changes. The mean ADAS-Noncog score for the original 27 subjects with AD was 4.4 (SD=3.5).¹ Specific items include:

1. Tearful
2. Appears/reports depressed mood
3. Concentration and distractibility
4. Uncooperative to testing
5. Delusions
6. Hallucinations
7. Pacing
8. Increased motor activity
9. Tremors
10. Increase or decrease in appetite

The cognitive subscale (ADAS-Cog) includes 11 tasks that are either a test to be completed by a subject or an assessment made by the test administrator about the subject, and which broadly assess the cognitive domains of memory, language, and praxis. The ADAS-Cog is scored from 0 to 70, and the mean ADAS-Cog score for the initial 27 subjects with AD was 19.3 (SD=12.1).¹ Specific tasks include:

1. Word Recall. A list of 10 words is read by the subject, and then the subject is asked to verbally recall as many of the words as possible. Three trials of reading and recalling are performed. The task score is the mean number of words not recalled across the three trials (range 0 to 10).¹

2. Naming Objects and Fingers. The subject is asked to name the fingers of their dominant hand as well as twelve objects, including: flower (plastic), bed (doll house furniture), whistle, pencil, rattle, mask, scissors, comb, wallet, harmonica, stethoscope, and tongs. The task score is calculated based on the number of fingers and objects correctly named, and ranges from 0 to 4.¹
3. Commands. The subject is asked to perform one to five step commands. For example, the two step command is to “Point to the ceiling, then to the floor.” The task score is from 0 to 5, based on the largest number of steps that are correctly performed (score is 0 if five step command is correctly performed).¹
4. Constructional Praxis. The subject is shown four geometric forms (circle, two overlapping rectangles, rhombus, cube) and asked to copy them on a piece of paper. The task is scored from 0 to 5 based on the number of correctly drawn forms.¹
5. Ideational Praxis. The subject is asked to pretend to send a letter to themselves. Scoring is based on difficulty of performing the five components of: fold letter, put letter in envelope, seal envelope, address envelope, and putting a stamp on the envelope (range 0 to 5).¹
6. Orientation. The subject is asked the date, month, year, day of the week, season, time of day, place, and person. The number of correct responses is the task score (range 0 to 8).¹
7. Word Recognition. The subject reads twelve words aloud, and then these twelve words are randomly shuffled with twelve new words, and the subject is asked whether they have previously seen each of the twenty-four words. Three trials are performed, and the task score is the mean number of correct responses across the three trials (range 0 to 12).¹
8. Language. After the administration of the Word Recall task (Q1) ten minutes of open-ended conversation occur between the test administrator and subject, before the remainder of the tasks are presented. These ten minutes of conversation are

used to assess language ability. Quality of speech is given a global rating by the administrator that ranges from 0 to 5.¹

9. Comprehension of Spoken Language. This task also relies on the ten minutes of open-ended conversation. The administrator provides an assessment of how well the subject can understand speech that ranges from 0 to 5.¹
10. Word Finding Difficulty. This task is also rated by the administrator during spontaneous speech to assess how much difficulty the subject has in finding desired words from 0 to 5.¹
11. Remembering Test Instructions. This task is a rating by the administrator from 1 to 5 according to the number of times that the subject needed to be reminded of instructions for the Word Recognition task.¹

Initially the two memory tasks (numbers 1 and 7) were viewed as a separate memory subscale and scored out of 22 points, with the remainder of the cognitive tasks scored out of 48 points (ADAS-Cog 9).¹ From here on “ADAS-Cog 11” refers to the full 11-item version of the ADAS-Cog.

3.1.2 Validation of the Alzheimer’s Disease Assessment Scale

Concurrent criterion validation was assessed in the original sample by correlating ADAS scores with previously well-established measures used to help assess disease severity. There were statistically significant correlations between the Sandoz Clinical Assessment-Geriatric and the full ADAS ($r=0.52$, $P<0.02$) as well as the ADAS-Cog 9 ($r=0.67$, $P<0.01$), but not the ADAS-Noncog ($r=0.25$, $P>0.10$).¹ There were statistically significant correlations between the Memory-Information Test and the full ADAS ($r=-0.67$, $P<0.001$), the ADAS-Cog ($r=-0.78$, $P<0.001$), and the ADAS-Noncog ($r=-0.42$, $P<0.02$).¹ There were also statistically significant correlations between the Dementia Rating Scale and the full ADAS ($r=0.64$, $P<0.001$), the ADAS-Cog ($r=0.48$, $P<0.01$), and the ADAS-Noncog ($r=0.46$, $P<0.01$).¹

Further concurrent criterion validation was performed in a separate study of 61 subjects with very mild, mild, moderate, or severe AD, and 52 subjects with NC.² The ADAS-Cog 11 and a modified ADAS-Noncog (nine items: tearfulness, depression, concentration, uncooperativeness, delusions, pacing, increased motor activity, tremors, appetite) were administered. The ADAS-Cog 11 correlated strongly with the Mini-Mental State Examination (MMSE; $r=-0.76$, $P<0.0001$), but there was a weaker correlation between the modified ADAS-Noncog and MMSE ($r=-0.39$, $P=0.0019$).²

3.1.3 Responsiveness to Baseline Discrimination

In the original ADAS development sample, point-biserial correlations were used to show that the group of subjects with AD had significantly higher scores on the ADAS-Cog 11 ($r=0.754$, $P<0.0001$) as well as on all individual ADAS-Cog 11 tasks (all $P<0.0001$) than the group of subjects with NC.¹ Subjects with AD also scored significantly worse on the ADAS-Noncog ($r=0.487$, $P<0.003$), and three individual ADAS-Noncog items.¹ Since then, several other studies with larger samples have also shown that the ADAS-Cog 11 is able to discriminate between the diagnostic categories of NC, Mild Cognitive Impairment (MCI), and AD at one point in time, and that the scores for subjects with NC are appropriately lower than those with MCI and subsequently AD.²⁻⁸ ADAS-Cog 11 scores have also been shown to discriminate between mild, moderate, and severe AD, but not between very mild and mild AD.² This remained true after removing the language tasks (Language, Comprehension of Spoken Language, Word Finding Difficulty) from the ADAS-Cog 11.² Another study including 485 subjects found statistically significant differences between ADAS-Cog 11 scores from subject groups with Clinical Dementia Rating (CDR) Scale scores of 0, 0.5, and 1.⁹

The ability of the ADAS-Cog 11 to act as a diagnostic instrument to classify subjects as having AD or not was tested by Zec et al. (1992), whereby two SD above the NC group mean was used as a cut-off for abnormal cognition.² Only two subjects with AD and one subject with NC were misclassified.² Good classification remained after removing the language tasks from the ADAS-Cog 11.²

3.1.4 Responsiveness to Disease Progression

Rosen et al. (1984) found a statistically significant worsening on total ADAS ($P=0.02$), ADAS-Cog 11 ($P=0.01$), and ADAS-Noncog ($P=0.03$) scores over a twelve-month period for ten subjects with AD, but not for ten of the subjects with NC (all $P>0.05$) that were used to develop the ADAS.¹ Eight of the subjects with AD showed a worsening on each individual task.¹

Evans et al. (2010) found statistically significant differences ($P<0.0005$ for all) between the magnitude of mean 12 month ADAS-Cog 11 change scores for subjects in different diagnostic categories, adjusted for baseline score, age, and gender, whereby the 99 subjects with AD changed the most (mean change=3.53 points, $SD=5.42$) compared to the 231 subjects with MCI (mean change=1.16 points, $SD=4.31$) and the 131 subjects with NC (mean change=-0.53 points, $SD=2.70$).¹⁰ Petersen et al. (2010) found statistically significant differences ($P<0.001$ for all) between mean ADAS-Cog 11 12 month change scores for 210 subjects with NC (mean change=-0.5 points, $SD=3.0$), 357 subjects with MCI (mean change=1.1 points, $SD=4.4$), and 161 subjects with AD (mean change=4.3 points, $SD=6.6$).⁴ It is important to note that the magnitudes of these changes are small, especially in MCI and NC groups. Other studies have found similar results. Steenland et al. (2014) found that ADAS-Cog 11 scores of 191 subjects with NC worsened by an average 7.5% over 3 years ($P=0.0007$) after adjusting for age, gender, race, education, and Apolipoprotein E (APOE) e4 allele presence.¹¹ Podhorna et al. (2016) did not perform statistical tests, but reported almost no change on the ADAS-Cog 11 in 382 subjects with MCI over 24 months (mean change=0.9 points, $SD=4.45$) and in 169 subjects with MCI over 36 months (mean change=1.9 points, $SD=5.45$).¹² For an 'enriched' subgroup of subjects with MCI who had cerebrospinal fluid (CSF) or APOEe4 allele biomarkers indicative of AD pathology, there was still only a 1.9 mean point change ($SD=4.92$) over 24 months ($n=206$) and 3.7 mean point change ($SD=6.21$) over 36 months ($n=89$).¹² In 97 subjects with mild AD there was a clinically relevant change over 12 months (mean change=3.5 points, $SD=5.59$) and for 40 subjects with mild AD over 24 months (mean change=8.3 points, $SD=8.96$).¹²

3.1.5 Minimum Clinically Relevant Change

Schrag et al. (2011) divided the scores of 358 subjects with mild AD into those who were rated by clinicians as having versus not having experienced clinically relevant worsening in the domains of memory, non-memory, general cognitive function, and functionality as assessed by the Functional Activities Questionnaire (FAQ) and CDR Scale.¹³ The range of ADAS-Cog score change corresponding to clinically relevant change separated by the four domains was 2.7 to 3.8 point increase over 6 months.¹³ For those judged not to have changed it was a mean 1.2 to 2.0 point increase over 6 months.¹³ All change scores were statistically significant ($P<0.05$).¹³ Based on these comparisons they determined that a 3 point or larger increase (worsening) on the ADAS-Cog 11 is a clinically relevant change.¹³

3.1.6 Summary

The ADAS was developed to assess the severity of cognitive and non-cognitive dysfunction in people with AD. The ADAS-Cog 11 and ADAS-Noncog are able to discriminate between groups of subjects with NC and AD, and the ADAS-Cog 11 can also discriminate MCI from NC and AD. The ADAS-Cog 11 has also been shown to be able to detect change over time in dementia and pre-dementia samples; however, the magnitude of the change detected in MCI and NC samples is very small. The ADAS-Noncog is not widely used and will not be reviewed further.

3.2 Assessment of the ADAS-Cog 11 in Pre-Dementia Populations

3.2.1 Ceiling Effects

Seven of the eleven ADAS-Cog 11 tasks demonstrate severe ceiling effects in MCI and NC samples (Table 1), whereby all or most subjects make zero errors on those tasks.^{5,14-19} A further two tasks show milder ceiling effects.^{14,15,17-19} Accordingly, 84% of errors made by subjects with NC and 71% of errors made by subjects with MCI occur on the two ADAS-Cog 11 tasks which do not demonstrate ceiling effects, Word Recall and Word

Recognition.⁵ Item Response Theory (IRT) analyses have also found that those two tasks have the most difficult ranking among all ADAS-Cog 11 tasks (please consult Section 2.3 for an overview of IRT).²⁰

Table 1 ADAS-Cog 11 Ceiling Effects.

| Study data | SIU-NAS ⁵ | Italian clinical trial ¹⁴ | Multiple pooled studies ¹⁵ | BCM-ADC ¹⁶ | ADCS ¹⁷ | ADNI ¹⁸ | ADNI ¹⁹ | ADNI ²¹ | ADNI ¹⁰ | |
|-------------|----------------------|--------------------------------------|---------------------------------------|-----------------------|--------------------|--------------------|--------------------|--------------------|--------------------|-----|
| Q1 | Green | | | | | | | | N/A | |
| Q2 | Red | | | | | | | N/A | N/A | |
| Q3 | Red | | | | | | | N/A | N/A | |
| Q4 | Orange | | Red | | | | Orange | | N/A | N/A |
| Q5 | Red | | | | | | | N/A | N/A | |
| Q6 | Orange | | Red | | | | Orange | | N/A | N/A |
| Q7 | Green | | | | | | | | N/A | N/A |
| Q8 | Red | | | | | | | N/A | N/A | |
| Q9 | Red | | | | | | | N/A | N/A | |
| Q10 | Red | | | | | | | N/A | N/A | |
| Q11 | Red | | | | | | | N/A | N/A | |
| Total Score | Green | | | N/A | Green | | | N/A | Green | |

Legend: Green=No ceiling effect, Orange=Mild ceiling effect, Red=Severe ceiling effect; ADCS=Alzheimer's Disease Cooperative Study, ADNI=Alzheimer's Disease Neuroimaging Initiative, BCM-ADC=Baylor College of Medicine Alzheimer's Disease Study, N/A=Not Available, Q1=Word Recall, Q2=Naming Objects and Fingers, Q3=Commands, Q4=Constructional Praxis, Q5=Ideational Praxis, Q6=Orientation, Q7=Word Recognition, Q8=Language, Q9=Comprehension of Spoken Language, Q10=Word Finding Difficulty, Q11=Remembering Test Instructions, SIU-NAS=Southern Illinois University Normal Aging Study. All studies included participants with Mild Cognitive Impairment except BCM-ADC, which included exclusively cognitively normal controls.

3.2.2 Information Content

An ideal outcome measure for MCI would have the information peaks of all item information curves generated by an IRT analysis of all individual outcome measure items in the range of cognitive ability expected to be seen in subjects with MCI. Ueckert et al. (2014) found that all but three ADAS-Cog 11 tasks have their information peak in a range that corresponds to levels of cognitive dysfunction more severe than would be expected in MCI, indicating that they are not optimally sensitive for use with MCI populations.¹⁵ Through summation of the information available for each subtask item (individual

response options), Ueckert et al. (2014) found that the most informative ADAS-Cog 11 tasks for assessing MCI levels of cognitive dysfunction were Word Recall, Orientation, Word Recognition, and Naming Objects and Fingers.¹⁵ Furthermore, an in-depth evaluation of the Word Recall Task has shown that the Pole response item has a higher recall probability than other response items for NC, MCI, and AD groups, suggesting that it is an abnormally easy item on the ADAS-Cog 11.²¹ These types of in-depth analyses for other individual ADAS-Cog 11 tasks have not been published.

3.2.3 Invariances

All individual ADAS-Cog 11 tasks as well as the total score have shown measurement invariance with respect to education and age in MCI samples.⁵ Measurement invariance to sex has also been found for the total ADAS-Cog 11 score in MCI samples.⁵ In samples with NC, the ADAS-Cog 11 total score also showed measurement invariance with respect to sex and education, but not for age.^{2,5,16} Age was significantly correlated with total ADAS-Cog 11 score as well as the Word Recall task in NC samples.^{2,5,16} For example, the ADAS-Cog 11 validation study that included 61 subjects with very mild, mild, moderate, or severe AD, and 52 subjects with NC recruited an additional 80 subjects with NC to assess whether age and education correlated with ADAS-Cog 11 scores.² They found ADAS-Cog 11 scores from subjects with NC were moderately correlated with age ($r=0.42$, $P=0.0018$), but non-significantly correlated with education ($r=-0.21$, $P=0.13$).² When categorized, age remained significantly correlated and education did not, suggesting that age but not education may influence ADAS-Cog 11 scores in NC samples.² Specifically, among the subjects with NC, those aged 7 to 13 or 60 to 89 years old performed significantly worse than those aged 14 to 59 years old, although the size of these differences was small.² In contrast, among subjects with AD correlations with age ($r=-0.08$, $P=0.56$) and education ($r=-0.06$, $P=0.66$) were both small and non-significant.² Altogether these results suggest that the only threat to measurement invariance is the age of subjects with NC.

3.2.4 Reliability

Significant variance in administration procedures and materials used for the ADAS-Cog 11 across clinical trials has been found, which threatens inter-observer, intra-observer, and test-retest reliability.²² Learning effects may also be a concern as Herholz et al. (2011) found a statistically significant decline in ADAS-Cog 11 scores in sample of subjects whom otherwise did not appear to be progressing in symptoms (stable NC) (ICC=0.47, 95%CI 0.32, 0.63).²³

3.2.5 Concurrent Criterion Validation

One study found ADAS-Cog 11 scores significantly correlated with MMSE scores in both NC (Spearman rho=-0.29, $P<0.001$) and MCI (Spearman rho=-0.66, $P<0.001$) samples, indicating agreement with another well-established assessment of overall cognitive ability.⁵ However, the only individual ADAS-Cog 11 tasks significantly correlated with the MMSE were Word Recognition in subjects with NC (Spearman rho=-0.26, $P<0.001$) and Word Recognition and Word Recall in subjects with MCI (Spearman rho range -0.36 to -0.49, $P<0.001$).⁵ Another study in 124 subjects with NC found the ADAS-Cog 11 was not significantly correlated with MMSE scores ($r=-0.13$, $P=0.16$).¹⁶

3.2.6 Responsiveness at the Item Level

Baseline discrimination. All ADAS-Cog 11 tasks have shown statistically significant differences between NC and MCI subgroups,^{4,5,17} and all but three tasks (Commands, Ideational Praxis, Language) have demonstrated significantly higher scores in AD than MCI subgroups.⁵ Furthermore, three of the ADAS-Cog 11 tasks (Word Recall, Word Recognition, Orientation) were found to detect a statistically significant difference between subjects with MCI and none versus one versus two APOEε4 alleles.²⁴

Disease progression. All individual ADAS-Cog 11 tasks have been found to have smaller Standardized Response Means (SRM)s than the ADAS-Cog 11 total score, where the three tasks demonstrating the largest SRM were Word Recall, Orientation, and Word Recognition.¹⁹ Groups of subjects with NC compared to MCI have been found to have

statistically significant different 12 month change scores on the Word Recall and Word Recognition tasks.⁴ The magnitude of 12 month and 24 month change scores for five ADAS-Cog 11 tasks were similar when comparing MCI and AD groups, while the six other tasks produced smaller change scores for the MCI group compared to the AD group.¹⁹

3.2.7 Performance of the ADAS-Cog 11 as an Outcome Measure in Pre-Dementia Studies

Forty-six studies were found which use the ADAS-Cog 11 as an outcome measure to assess whether there is an association between an exposure or intervention and cognitive ability. This thesis takes a very rudimentary approach to assessing the performance of the ADAS-Cog 11 in these studies, and merely examined whether or not the ADAS-Cog 11 produced statistically significant results for these associations. It is assumed that all studies in this portion of the literature review have an adequately developed theoretical framework to reasonably expect that a difference in cognitive ability between exposure groups should exist, either at baseline or over time, even though some of the exposures may actually be ineffectual towards cognitive ability. Results from other outcome measures used to assess the same association as the ADAS-Cog 11 in each study may be used as a sort of proxy for whether the ADAS-Cog 11 is capturing associations which truly do exist (other measures statistically significant). Please note that due to publication bias towards positive results, the results presented here may overestimate the proportion of statistically significant associations detected by the ADAS-Cog 11 in pre-dementia study samples.

Responsiveness to group-level between-person differences in observed level of disease severity based on exposure status. Twenty-two studies assessed cross-sectional associations between exposure status and ADAS-Cog 11 scores in older adults with pre-dementia levels of impairment, as summarized below in Table 2 for NC, Table 3 for MCI, and Table 4 for mixed NC and MCI samples.^{4,11,25-44} Within these studies there were twenty statistically significant associations between exposure status and ADAS-Cog 11 scores (green highlight). There were sixteen non-statistically significant associations

found between the ADAS-Cog 11 and an exposure, where any other cognitive or brain imaging outcome measures used to assess the same association also produced non-statistically significant results (orange highlight). The ADAS-Cog 11 failed to produce a statistically significant result for eight associations which were statistically significant for at least one other cognitive or brain imaging outcome measure (red highlight).

Table 2 Normal Cognition Samples: ADAS-Cog 11 responsiveness to group-level between-person differences in observed level of disease severity based on exposure status.

| First Author & Publication Year | Exposure (continuous variable unless otherwise specified) | Association between exposure and ADAS-Cog 11 [Effect estimate, P-Value, (n)] | Other statistically significant outcome measures | Other statistically non-significant outcome measures | Factors controlled for (None if blank) |
|---------------------------------|--|---|--|--|---|
| Crane 2012 ²⁵ | Hippocampal Volume | $\beta=1.03$, P>0.05, (225) | | ADNI-Mem, RAVLT, ADAS-Cog13, ADAS-Rasch, ADAS-Tree, MMSE, CDR-SB | Age, education, gender, APOEe4 allele, intracranial volume |
| Crane 2012 ²⁵ | Parahippocampal Thickness | $\beta=1.30$, P>0.05, (225) | | ADNI-Mem, RAVLT, ADAS-Cog13, ADAS-Rasch, ADAS-Tree, MMSE, CDR-SB | Age, education, gender, APOEe4 allele, intracranial volume |
| Crane 2012 ²⁵ | Entorhinal Thickness | $\beta=0.27$, P>0.05, (225) | | ADNI-Mem, RAVLT, ADAS-Cog13, ADAS-Rasch, ADAS-Tree, MMSE, CDR-SB | Age, education, gender, APOEe4 allele, intracranial volume |
| Crane 2012 ²⁵ | Fusiform Thickness | $\beta=-0.17$, P>0.05, (225) | ADNI-Mem, RAVLT | ADAS-Cog13, ADAS-Rasch, ADAS-Tree, MMSE, CDR-SB | Age, education, gender, APOEe4 allele, intracranial volume |
| Daiello 2015 ²⁶ | Fish Oil Supplement vs. None | $\beta=-7.01$, P<0.01, (229) | MMSE | | Age, gender, education, race, CVD risk score, APOEe4 allele, ChEI use |
| Doraiswamy 2012 ²⁷ | A β positive vs. negative | P=0.17, (69) | DSS, WMS immediate & delayed recall | MMSE, CDR-SB, verbal fluency (animals & vegetables) | |

| | | | | | |
|-------------------------------|---|-------------------------|---|--|--|
| Doraiswamy 2014 ²⁸ | A β positive vs. negative | P=0.20, (67) | DSS, WMS immediate recall | MMSE, WMS delayed recall, CDR-SB, verbal fluency (animals & vegetables) | |
| Landau 2012 ²⁹ | Brain Glucose Metabolism | P>0.05, (126) | | | |
| Landau 2012 ²⁹ | A β | rho=0.17, P=0.06, (126) | | | |
| Petersen 2010 ⁴ | A β | r=-0.21, P<0.05, (229) | | | |
| Steenland 2014 ¹¹ | Future Conversion to MCI or AD vs. No Future Conversion | P=0.09, (191) | RAVLT trial 5 & short recall, WMS immediate & delayed logical memory, BNT | Mini-Cog, MMSE, ANART, Category (animal) fluency, TMTA&B, brain volume measures (whole brain, ventricle, left hippocampus, right hippocampus). | |

Legend: See List of Abbreviations (pages xii to xvi). Effect estimates include regression coefficients (β), Spearman rank correlation coefficients (rho), and Pearson correlation coefficients (r). If a statistical test was performed without an effect estimate reported, only the P-value is shown. Highlighting refers to results of associations tested using the ADAS-Cog as an outcome measure: Green=statistically significant result, suggesting responsiveness; Orange=non-statistically significant result where no other cognitive or neuroimaging outcome measure found a statistically significant association, suggesting unknown responsiveness; Red=non-statistically significant result where at least one other cognitive or neuroimaging outcome measure detected a statistically significant result, suggesting poor responsiveness of the ADAS-Cog.

Table 3 Mild Cognitive Impairment Samples: ADAS-Cog 11 responsiveness to group-level between-person differences in observed level of disease severity based on exposure status

| First Author & Publication Year | Exposure (continuous variable unless otherwise specified) | Association between exposure and ADAS-Cog 11 [Effect estimate, P-Value, (n)] | Other statistically significant outcome measures | Other statistically non-significant outcome measures | Factors controlled for (None if blank) |
|---------------------------------|--|---|--|--|---|
| | | | | | |

| | | | | | |
|-------------------------------|---------------------------------|-------------------------------------|--|---|---|
| Crane 2012 ²⁵ | Hippocampal Volume | $\beta=6.32$, P<0.05, (394) | ADNI-Mem, RAVLT, ADAS-Cog13, ADAS-Rasch, ADAS-Tree, MMSE, CDR-SB | | Age, education, gender, APOEe4 allele, intracranial volume |
| Crane 2012 ²⁵ | Parahippocampal Thickness | $\beta=1.63$, P>0.05, (394) | ADNI-Mem, RAVLT, ADAS-Tree, CDR-SB | ADAS-13, ADAS-Rasch, MMSE | Age, education, gender, APOEe4 allele, intracranial volume |
| Crane 2012 ²⁵ | Entorhinal Thickness | $\beta= 7.68$, P<0.05, (394) | ADNI-Mem, RAVLT, ADAS-Cog13, ADAS-Rasch, ADAS-Tree, MMSE, CDR-SB | | Age, education, gender, APOEe4 allele, intracranial volume |
| Crane 2012 ²⁵ | Fusiform Thickness | $\beta= 4.46$, P<0.05, (394) | ADNI-Mem, RAVLT, ADAS-Cog13, ADAS-Rasch, ADAS-Tree, MMSE, CDR-SB | | Age, education, gender, APOEe4 allele, intracranial volume |
| Cronk 2010 ³⁰ | Body Mass Index | No statistical test, (286) | | MMSE, global cognition score | Age, sex, education |
| Daiello 2015 ²⁶ | Fish Oil Supplement vs. None | $\beta=-3.29$, P=0.20, (397) | | MMSE | Age, gender, education, race, CVD risk score, APOEe4 allele, ChEI |
| Doraiswamy 2012 ²⁷ | A β positive vs. negative | P=0.06, (51) | | MMSE, CDR-SB, DSS, verbal fluency (vegetables & animals), WMS immediate & delayed recall | |
| Doraiswamy 2014 ²⁸ | A β positive vs. negative | P=0.10, (47) | | MMSE, CDR-SB, DSS, verbal fluency (animal & vegetable), WMS immediate & delayed recall | |
| Irizarry 2009 ³¹ | Urate Quintiles | P=0.65, (747) | | | |

| | | | | | |
|----------------------------|--|-------------------------------------|---|--|---|
| Kennedy 2012 ³² | $A\beta \leq$ vs. > 192 pG/mL | P=0.002, (198) | WMS delayed logical memory, RAVLT delayed | MMSE, CDR-SB | |
| Kennedy 2012 ³² | t-Tau/ $A\beta$ $>$ vs. ≤ 0.39 | P<0.001, (198) | WMS delayed logical memory delay, RAVLT delay | MMSE, CDR-SB | |
| Kennedy 2014 ³³ | APOE4 allele Carrier vs. Non-Carrier | P<0.001, (1192) | CDR-SB, MMSE | | |
| Kennedy 2016 ³⁴ | APOE4 allele Carrier vs. Non-Carrier | P<0.001, (1171) | | | |
| Landau 2012 ²⁹ | $A\beta$ | $\rho=0.24$, P=0.002, (162 early) | | | |
| Landau 2012 ²⁹ | $A\beta$ | $\rho=0.29$, P=0.007, (85 late) | | | |
| Landau 2012 ²⁹ | Brain Glucose Metabolism | $\rho=-0.25$, P=0.001, (162 early) | | | |
| Landau 2012 ²⁹ | Brain Glucose Metabolism | $\rho=-0.32$, P=0.003, (85 late) | | | |
| Mackin 2013 ³⁵ | Subsyndromal Symptoms of Depression vs. None | P=0.10, (405) | White matter lesion | MMSE | |
| McGough 2013 ³⁶ | Gait Velocity | $\beta=-0.19$, P=0.008, (201) | TMT A&B, WMS logical memory, word recall | | Age, sex, musculoskeletal comorbidity, depression symptoms |
| McGough 2013 ³⁶ | Physical Activity | $\beta=-0.10$, P=0.18, (201) | TMT B, Word recall | TMT A, WMS logical memory I | Age, sex, depressive symptoms, musculoskeletal comorbidity |
| McGough 2013 ³⁶ | Grip Strength | $\beta=-0.05$, P=0.40, (201) | TMT A | TMT B, Word Recall, WMS logical memory I | Age, sex, BMI, depressive symptoms, musculoskeletal comorbidity |

| | | | | | |
|------------------------------|---|-------------------------------|--|--|--|
| Petersen 2010 ⁴ | Future Progression to AD at 1 year vs. No Progression to AD | P<0.001, (398) | CDR-SB, MMSE, ADAS-Cog without word list, recall, and recognition items, RAVLT, TMT A&B, Category fluency (animal & vegetable), Number cancellation, BNT, Digit backwards, Clock drawing | CDR | |
| Petersen 2010 ⁴ | A β | r=-0.22, P<0.05, (398) | | | |
| Portelius 2015 ³⁷ | CSF Neurogranin Quartiles | No statistical test, (173) | | MMSE | |
| Rozzini 2006 ³⁸ | Progressive vs. Stable MCI | P=0.05, (74) | | | |
| Rozzini 2008 ³⁹ | Extrapyramidal Signs vs. None | P=0.03, (160) | ADAS-Cog without memory tasks | MMSE, CDR, ADAS-Cog memory tasks, Short Story (Novelli), Rey's figure copy, Phonologic verbal fluency, Semantic verbal fluency | |
| Schneider 2011 ⁴⁰ | ChEI vs. ChEI and Memantine Hydrochloride vs. Neither | P<0.001, (392) | CDR-SB | MMSE | |
| Toledo 2014 ⁴¹ | CSF levels Complement 3 | β =-0.061, P=1.0, (187) | | MMSE, memory & EF summary scores | Age, gender, education, APOEe4 allele, t-Tau/A β |
| Toledo 2014 ⁴¹ | CSF levels Factor H | β =-0.077, P=1.0, (187) | | MMSE, memory & EF summary scores | Age, gender, education, APOEe4 allele, t-Tau/A β |
| Toledo 2014 ⁴¹ | Complement 3/Factor H | β =0.076, P=1.0, (187) | | MMSE, memory & EF summary scores | Age, gender, education, APOEe4 allele, t-Tau/A β |

| | | | | | |
|------------------------------|---|----------------|---|--|--|
| Whitehair 2010 ⁴² | APOEε4 allele Carriers vs. Non-Carriers | P<0.001, (516) | ADAS-Cog13, NYU paragraph recall immediate and delayed recall, MMSE, CDR-SB, Clock drawing, Category fluency, Number cancellation target hits, SDMT | Digit backwards task, BNT, Number cancellation target errors, Maze tracing | |
|------------------------------|---|----------------|---|--|--|

Legend: See List of Abbreviations (pages xii to xvi). Effect estimates include regression coefficients (β), Spearman rank correlation coefficients (ρ), and Pearson correlation coefficients (r). If a statistical test was performed without an effect estimate reported, only the P-value is shown. Highlighting refers to results of associations tested using the ADAS-Cog as an outcome measure: Green=statistically significant result, suggesting responsiveness; Orange=non-statistically significant result where no other cognitive or neuroimaging outcome measure found a statistically significant association, suggesting unknown responsiveness; Red=non-statistically significant result where at least one other cognitive or neuroimaging outcome measure detected a statistically significant result, suggesting poor responsiveness of the ADAS-Cog.

Table 4 Mixed Normal Cognition and Mild Cognitive Impairment Samples: ADAS-Cog 11 responsiveness to group-level between-person differences in observed level of disease severity based on exposure status.

| First Author & Publication Year | Exposure (continuous variable unless otherwise specified) | Association between exposure and ADAS-Cog 11 [Effect Estimate, P-Value, (n)] | Other statistically significant outcome measures | Other statistically non-significant outcome measures | Factors controlled for (None if blank) |
|---------------------------------|--|---|--|--|---|
| Betterman 2012 ⁴³ | Lipid Lowering Medication vs. None | P=0.81, (3069) | | 3MSE | |
| Pernecky 2006 ⁴⁴ | Activities of Daily Living | r=-0.46, P<0.01 (75) | | | |

Legend: See List of Abbreviations (pages xii to xvi). Effect estimates include a Pearson correlation coefficient (r). If a statistical test was performed without an effect estimate

reported, only the P-value is shown. Highlighting refers to results of associations tested using the ADAS-Cog as an outcome measure: Green=statistically significant result, suggesting responsiveness; Orange=non-statistically significant result where no other cognitive or neuroimaging outcome measure found a statistically significant association, suggesting unknown responsiveness; Red=non-statistically significant result where at least one other cognitive or neuroimaging outcome measure detected a statistically significant result, suggesting poor responsiveness of the ADAS-Cog.

Responsiveness to group-level between-person differences of within-person observed change in those estimated to be different based on baseline exposure status. Twenty-two studies were found which tested for an association between baseline exposure status and change in ADAS-Cog 11 scores over a follow-up period, as summarized below in Table 5 for NC and Table 6 for MCI samples.^{4,10,11,14,23,25,27-29,32,33,35,38-41,43-48} Among these studies there were forty-three statistically significant associations between baseline exposure status and ADAS-Cog 11 scores over time (green highlight). There were twenty-one non-statistically significant associations between baseline exposure and ADAS-Cog 11 scores over time, whereby any other cognitive or brain imaging outcome measures also produced non-statistically significant results (orange highlight). The ADAS-Cog 11 produced a further three non-statistically significant results for associations found to be statistically significant by at least one other cognitive or brain imaging outcome measure (red highlight).

Table 5 Normal Cognition Samples: Responsiveness to group-level between-person differences of within-person observed change in those estimated to be different based on baseline exposure status.

| First Author & Publication Year | Exposure (continuous variable unless otherwise specified) | Association between exposure and change in ADAS-Cog 11 [Effect Estimate, P-Value, (n, years of follow-up)] | Other statistically significant outcome measures | Other statistically non-significant outcome measures | Factors controlled for (None if blank) |
|--|---|--|---|---|--|
| | | | | | |

| | | | | | |
|-------------------------------|-------------------------------------|---|---|--|---|
| Bettermann 2012 ⁴³ | Lipid Lowering Medications vs. None | P=0.04, (2578, 3) | 3MSE | | Age, sex, race, education, clinic, treatment group, MCI, APOEe4 allele, CHD, stroke |
| Crane 2012 ²⁵ | AD CSF Signature vs. No Signature | Z=-1.96 (reversed score), P<0.05, (112, 3) | ADAS-Cog 13, CDR-SB | ADNI-Mem, RAVLT, ADAS-Rasch, ADAS-Tree, MMSE | Age, education, sex, APOEe4 allele |
| Doraiswamy 2012 ²⁷ | A β Positive vs. Negative | P=0.005, (67, 1.5) | CDR-SB | MMSE, DSS, Verbal Fluency (animals & vegetables), WMS delayed & immediate recall | Age, psychometric assessment |
| Doraiswamy 2012 ²⁷ | Florbetapir SUVr | P=0.095, (67, 1.5) | | MMSE, CDR-SB, DSS, Verbal fluency (animals & vegetables), WMS delayed & immediate recall | Age, psychometric assessment |
| Doraiswamy 2014 ²⁸ | A β Positive vs. Negative | P=0.001, (67, 3) | CDR-SB, DSS, Verbal fluency (vegetable) | Verbal fluency (animal), WMS logical & immediate recall, MMSE | Age, cognitive function assessment |
| Landau 2012 ²⁹ | A β Positive vs. Negative | β =0.43, P<0.001, (76, 4) | | | Age, sex, education |
| Landau 2012 ²⁹ | Brain Hypometabolism | P>0.05, (76, 4) | | | Age, sex, education |
| Lo 2011 ⁴⁵ | A β | P>0.05, (36, 3) | | | Age, baseline biomarker |
| Lo 2011 ⁴⁵ | Brain Glucose Metabolism | P>0.05, (104, 3) | | | Age, baseline biomarker value |
| Lo 2011 ⁴⁵ | Hippocampal Volume | P>0.05, (228, 3) | | | Age, baseline biomarker value |
| Petersen 2010 ⁴ | A β | r=-0.23, P<0.05, (229, 1) | | | |
| Steenland 2014 ¹¹ | Age | β =0.15 (log[ADAS-Cog]), P=0.003, (191, 3.1) | | | Gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | Male | β =0.19 (log[ADAS-Cog]), P=0.0009, (191, 3.1) | | | Age, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | Race, white | β =0.05 (log[ADAS-Cog]), P=0.65, (191, 3.1) | | | Age, gender, education, APOEe4 allele, time |

| | | | | | |
|------------------------------|---|---|--|--|--|
| Steenland 2014 ¹¹ | Education | $\beta=-0.03$ (log[ADAS-Cog]), P=0.002, (191, 3.1) | | | Age, gender, race, APOEe4 allele, time |
| Steenland 2014 ¹¹ | APOEe4 allele | $\beta=0.14$ (log[ADAS-Cog]), P=0.03, (191, 3.1) | | | Age, gender, race, education, time |
| Steenland 2014 ¹¹ | Category (Animal) Fluency | P=0.05, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | Whole Brain Volume | P=0.02, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | Hippocampal Volume | P=0.02 (left) P=0.008 (right), (186, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | t-Tau | P=0.04, (188, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | p-Tau | P=0.006, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | A β | P=0.0007, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | t-Tau/A β | P=0.01, (188, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | p-Tau/A β | P=0.003, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | Mini-cog | P>0.05, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | MMSE | P>0.05, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | ANART | P>0.05, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | RAVLT trial 5 | P>0.05, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | RAVLT short recall | P>0.05, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | TMT A or B | P>0.05, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |
| Steenland 2014 ¹¹ | WMS Logical Memory (immediate or delayed) | P>0.05, (191, 3.1) | | | Age, gender, race, education, APOEe4 allele, time |

| | | | | | |
|------------------------------|---------------------------|----------------------------|--|------|--|
| Steenland 2014 ¹¹ | Boston Naming Test | P>0.05, (191, 3.1) | | | Age, gender, race, education, APOEε4 allele, time |
| Steenland 2014 ¹¹ | Ventricle Volume | P>0.05, (191, 3.1) | | | Age, gender, race, education, APOEε4 allele, time |
| Ye 2016 ⁴⁶ | Serum Uric Acid (Females) | β=0.10, P=0.02, (137, 2.9) | | MMSE | Age, sex, education, BMI, race, APOEε4 allele, cardiovascular risk factors, study site |
| Ye 2016 ⁴⁶ | Serum Uric Acid (Males) | β=0.01, P=0.88, (134, 2.9) | | MMSE | Age, sex, education, BMI, race, APOEε4 allele, CVD risk factors, study site |

Legend: See List of Abbreviations (pages xii to xvi). Effect estimates include regression coefficients (β), Z-scores from mixed effects models (Z), and Pearson correlation coefficients (r). If a statistical test was performed without an effect estimate reported, only the P-value is shown. Highlighting refers to results of associations tested using the ADAS-Cog as an outcome measure: Green=statistically significant result, suggesting responsiveness; Orange=non-statistically significant result where no other cognitive or neuroimaging outcome measure found a statistically significant association, suggesting unknown responsiveness; Red=non-statistically significant result where at least one other cognitive or neuroimaging outcome measure detected a statistically significant result, suggesting poor responsiveness of the ADAS-Cog.

Table 6 Mild Cognitive Impairment Samples: Responsiveness to group-level between-person differences of within-person observed change in those estimated to be different based on baseline exposure status.

| First Author & Publication Year | Exposure (continuous variable unless otherwise specified) | Association between exposure and change in ADAS-Cog 11 [Effect Estimate, P-Value, (n, years of follow-up)] | Other statistically significant outcome measures | Other statistically non-significant outcome measures | Factors controlled for (None if blank) |
|---------------------------------|--|--|--|--|---|
| Bettermann 2012 ⁴³ | Lipid Lowering Medications vs. None | P>0.05, (491, 3) | | 3MSE | Age, sex, race, education, clinic, treatment group, MCI, APOEε4 allele, CHD, stroke |

| | | | | | |
|-------------------------------|-------------------------------------|---|--|---|--|
| Cardinali 2012 ⁴⁷ | Melatonin vs. None | $Z_{MW}=-5.73$, $P<0.001$, (96, 5) | MMSE, Mattis' score, DSS, TMT A&B, RAVLT | | |
| Crane 2012 ²⁵ | AD CSF Signature vs. No Signature | $Z=-4.39$, $P<0.05$, (193, 3) | ADNI-Mem, RAVLT, ADAS-Cog13, ADAS-Rasch, ADAS-Tree, MMSE, CDR-SB | | Age, education, sex, APOEε4 allele |
| Cronk 2010 ³⁰ | BMI | Wald $X^2=6.7$, $P=0.02$. (286, 1) | MMSE, Global composite | CDR-SB | Age, education, sex |
| Doraiswamy 2012 ²⁷ | Aβ positive vs. negative | $P=0.001$, (46, 1.5) | MMSE, CDR-SB, DSS, Verbal fluency (vegetables), WMS delayed & immediate recall | Verbal fluency (animals) | Age, baseline psychometric score |
| Doraiswamy 2012 ²⁷ | Florbetapir SUVr | $r=0.41$, $P=0.006$, (46, 1.5) | MMSE, CDR-SB, DSS, WMS immediate recall | Verbal fluency (animal & vegetable), WMS delayed recall | Age, baseline psychometric score |
| Doraiswamy 2014 ²⁸ | Aβ positive vs. negative | $P=0.001$, (46, 3) | CDR-SB, DSS, verbal fluency (vegetable), MMSE | Verbal fluency (animal), WMS logical & immediate memory | Age, baseline cognitive function scores |
| Evans 2010 ¹⁰ | Brain Atrophy Rates | $P<0.0001$, (231, 1) | MMSE, TMT B | TMT A | Baseline brain volume, neuropsychological score, age, gender |
| Evans 2010 ¹⁰ | Ventricular Expansion | $P<0.0005$, (231, 1) | MMSE, TMT B | TMT A | Baseline brain volume, neuropsychological score, age, gender |
| Furio 2007 ⁴⁸ | Melatonin vs. None | $Z_{MW}=-5.55$, $P=0.001$, (50, 1.5) | MMSE, Mattis' score, TMT A&B, RAVLT | DSS | |
| Herholz 2011 ²³ | Progressive vs. Non-Progressive MCI | Cohen $D=0.30$, $P>0.05$, (94, 1) | PET measure | | |

| | | | | | |
|------------------------------|--|--|---|--|--|
| Herholz 2011 ²³ | Progressive vs. Non-Progressive MCI | Cohen D=0.60, P=0.006 (94, 2) | PET measure | | |
| Irizarry 2009 ²⁹ | Plasma Urate | P>0.05, (747, 3) | | | Age, sex, BMI, APOEε4 allele, current smoking, history of alcohol abuse, CVD, hypertension, use of nonsteroidal anti-inflammatory drugs and thiazide diuretics |
| Kennedy 2016 ³⁴ | APOEε4 allele Present vs. Absent | P<0.001, (1171, 3) | | | |
| Landau 2012 ²⁹ | Aβ Positive vs. Negative | β=0.83, P=0.004, (81, 4) | | | Age, sex, education |
| Landau 2012 ²⁹ | Brain Glucose Hypometabolism Positive vs. Negative | β=1.48, P<0.001, (81, 4) | | | Age, sex, education |
| Lo 2011 ⁴⁵ | Aβ | P<0.05, (54, 3) | | | Age, baseline biomarker value |
| Lo 2011 ⁴⁵ | Brain Glucose Metabolism | P<0.001, (203, 3) | | | Age, baseline biomarker value |
| Lo 2011 ⁴⁵ | Hippocampal Volume | P<0.001, (390, 3) | | | Age, baseline biomarker value |
| Mackin 2013 ³⁵ | Subsyndromal Symptoms of Depression vs. None | β _{GEE} =0.51, P=0.28, (405, 2) | | | APOEε4 allele |
| Petersen 2010 ⁴ | Aβ | r=-0.29, P<0.05, (398, 1) | | | |
| Portelius 2015 ³⁷ | CSF Neurogranin Quartiles | β=0.002, P=0.0002, (173, 9) | MMSE, hippocampal volume, cortical glucose metabolism | | Age, sex, education |
| Schneider 2011 ⁴⁰ | ChEIs vs. None | β=0.78, P=0.03, (392, 2) | MMSE, CDR-SB | | Age, APOEε4 allele, education, baseline ADAS-Cog or CDR-SB |
| Schneider 2011 ⁴⁰ | ChEI and Memantine Hydrochloride vs. None | β=0.86, P=0.14, (251, 2) | MMSE, CDR-SB | | Age, APOEε4 allele, education, baseline ADAS-Cog or CDR-SB |

| | | | | | |
|------------------------------|--|-------------------------------------|--|--|---|
| Schneider 2011 ⁴⁰ | ChEIs and Memantine Hydrochloride vs. ChEIs only | P>0.05, (177, 1.5) | MMSE, CDR-SB | | |
| Toledo 2014 ⁴¹ | CSF levels Complement 3 | β =-0.12, P=0.04, (160, 1.5) | | MMSE, Memory assessment, EF assessment | Age, gender, APOEe4 allele, education, t-Tau/A β |
| Toledo 2014 ⁴¹ | CSF levels Factor H | β =-0.08, P=0.04, (160, 1.5) | | MMSE, Memory assessment, EF assessment | Age, gender, APOEe4 allele, education, t-Tau/A β |
| Toledo 2014 ⁴¹ | Complement 3/Factor H | β =-0.18, P=0.06, (160, 1.5) | | MMSE, memory assessment, EF assessment | Age, gender, APOEe4 allele, education, t-Tau/A β |
| Whitehair 2010 ⁴² | APOEe4 Allele Present vs. Absent | P<0.001, (516, 3) | ADAS-Cog13, Delayed work list recall, MMSE, Digit backwards, BNT, Clock drawing, Category fluency, New York University immediate & delayed paragraph recall, Number cancellation target hits, SDMT, CDR-SB | Number cancellation task target errors, maze tracing | Age, sex, education, baseline CDR-SB |
| Ye 2016 ⁴⁶ | Serum Uric Acid | β =-0.5, P<0.001, (244, 2.9) | MMSE | | Age, sex, education, BMI, race, APOEe4 allele, CVD risk factors, study site |
| Ye 2016 ⁴⁶ | Serum Uric Acid | β =-0.001, P=0.99, (352, 2.9) | | MMSE | Age, sex, education, BMI, race, APOEe4 allele, CVD risk factors, study site |

| | | | | | |
|-------------------------------|---|----------------------|--------------------------|--|--|
| Zanotta 2014 ¹⁴ | Phytotherapeutic Compound plus Phosphatidylserine and Vitamin E vs. Placebo | P<0.001, (102, 0.16) | Clock drawing test, MMSE | | |
|-------------------------------|---|----------------------|--------------------------|--|--|

Legend: See List of Abbreviations (pages xii to xvi). Effect estimates include regression coefficients (β), parameters from generalized estimating equations (β_{GEE}), Pearson correlation coefficients (r), and Z-scores from mixed effects models (Z) or Mann-Whitney U tests (Z_{MW}). If a statistical test was performed without an effect estimate reported, only the P-value is shown. Highlighting refers to results of associations tested using the ADAS-Cog as an outcome measure: Green=statistically significant result, suggesting responsiveness; Orange=non-statistically significant result where no other cognitive or neuroimaging outcome measure found a statistically significant association, suggesting unknown responsiveness; Red=non-statistically significant result where at least one other cognitive or neuroimaging outcome measure detected a statistically significant result, suggesting poor responsiveness of the ADAS-Cog.

Responsiveness to treatment effects. Seventeen clinical trials using the ADAS-Cog 11 as an outcome measure in pre-dementia samples were found, and are summarized in Table 7, below.⁴⁹⁻⁶⁵ The ADAS-Cog 11 was able to detect seven statistically significant treatment effects (green highlight). The ADAS-Cog 11 did not find a significant effect for eleven interventions (orange or red highlight), four of which demonstrated a treatment effect for at least one other outcome measure (red highlight). Note that only results from the final time point of each study are presented, and subgroup analyses are only presented when the primary analyses did not include the ADAS-Cog 11 in a sample composed completely of older adults with pre-dementia levels of disease severity.

Table 7 ADAS-Cog 11 Responsiveness to Treatment Effects in Pre-Dementia

Clinical Trials

| First Author & Publication Year | Treatment vs. Placebo (unless otherwise specified) | Treatment effect [P-Value, (n, years of follow-up)] | Other outcome measures (* if sig effect) | ADAS-Cog as primary outcome? | Factors controlled for (None if blank) |
|--|--|---|--|-------------------------------------|--|
|--|--|---|--|-------------------------------------|--|

| | | | | | |
|--------------------------------------|---|---|--|-----|-----------------------------------|
| De Beaumont 2016 ⁴⁹ | Donepezil | Stratified by genotype BCHE-K*: P<0.01; BCHE-K wild type: P>0.05; APOEε4+: P>0.05; APOEε4-: P>0.05, (408, 3) | | Yes | Age, sex, baseline ADAS-Cog |
| Buschert 2011 ⁵⁰ | Multicomponent Cognitive Group Intervention | P=0.02, (22, 0.5) | MMSE, RBANS (story memory, recall*), TMT A&B, MADRS*, QoL-AD | Yes | Age, education |
| Buschert 2012 ⁵¹ | Multicomponent Cognitive Group Intervention | P=0.04, (24, 2.3) | RBANS (immediate*, delayed), TMT A&B, MMSE, MADRS, QoL-AD | Yes | |
| Chiu 2008 ⁵² | Omega-3 Polyunsaturated Fatty Acids | P=0.03, (23, 0.5) | CIBIC-Plus | Yes | Age, gender, education |
| Dubois 2012 ⁵³ | Pro-Cholinergic Drug | P=0.37, (241, 0.46) | CDR-SB, CVLT(free immediate recall Monday list & Tuesday list*, short delay free and cued recall, long delay free and cued recall), Fluency test, TMT A&B, DSST, Global improvement evaluated by investigator & patient, ADL, AI | Yes | Country |
| Forster 2011 ⁵⁴ | Cognitive Intervention | P=0.045, (21, 0.5) | MMSE* | Yes | Education, age |
| Kile 2015 ⁵⁵ | Immunoglobulin | P=0.03, (49, 2) | Annualized percent change in ventricular volume, MMSE*, CDR-SB | No | |
| Lin 2014 ⁵⁶ | Sodium Benzoate | P=0.23, (31, 0.46) | CIBIC-Plus, Cognitive composite | Yes | |
| Luchsinger 2016 ⁵⁷ | Metaformin | P=0.34, (80, 1) | SRT*, Glucose uptake in posterior cingulate-precuneus, Aβ | Yes | |
| Miao 2012 ⁵⁸ | Chinese Herbal Medicine vs. Donepezil | P=0.11, (72, 1) | MMSE, ADL, Syndrome Differentiation Scale | Yes | |

| | | | | | |
|-----------------------------------|--|----------------------|---|-----|---------------------|
| Petersen 2005 ⁵⁹ | Donepezil and Vitamin E | P>0.05, (769, 3) | MMSE, CDR, CDR-SB, ADL, Global deterioration scale, Neuropsychological battery, Time to development of AD | No | |
| De Gobbi Porto 2015 ⁶⁰ | Aerobic Training | P<0.001, (40, 0.46) | Neuropsychological battery* | Yes | |
| Reuter 2012 ⁶¹ | Cognitive Training+Transfer Training+Psychomotor Training vs. CT+TT vs. CT | P<0.001, (223, 0.58) | SCOPA-Cog*, BADS (zoo*, instruction*, 6 elements*), Paced auditory serial addition test* | Yes | Age, sex, education |
| Singh 2014 ⁶² | Computerized Multidomain Cognitive Training | P=0.69, (51, 1.5) | Standardized global cognition score, WAIS-III similarities & matrices, Category fluency, COWAT, Memory function, BVRT, Immediate & Delayed & Domain memory scores, SDMT, ADL | Yes | |
| Singh 2014 ⁶² | High Intensity Progressive Resistance Training | P=0.08, (49, 1.5) | Standardized global cognition score, WAIS-III Similarities & matrices*, Category fluency, COWAT, Memory function, BVRT, Immediate & Delayed & Domain Memory scores, SDMT, ADL | Yes | |
| Snitz 2009 ⁶³ | G Biloba Extract | P=0.97, (3069, 7.3) | 3MSE, Tests of memory, attention, visual-spatial construction, language, and EF | Yes | |
| Suzuki 2013 ⁶⁴ | Multicomponent Exercise with Multitask Conditions vs. Educational Classes | P=0.16, (100, 0.5) | MMSE, WMS-logical memory I & II, Volume of medial temporal areas including the entorhinal cortex, Whole brain cortices | No | |
| Thal 2005 ⁶⁵ | Rofecoxib | P>0.05, (1457, 4) | % patients convert to AD*, CDR-SB, SRT, MMSE | No | |

Legend: See List of Abbreviations (pages xii to xvi). P-values refer to statistical tests performed to assess whether a treatment effect was present. Highlighting refers to results of associations tested using the ADAS-Cog as an outcome measure: Green=statistically significant result, suggesting responsiveness; Orange=non-statistically significant result where no other outcome measure found a statistically significant association, suggesting

unknown responsiveness; Red=non-statistically significant result where at least one other outcome measure detected a statistically significant result, suggesting poor responsiveness of the ADAS-Cog.

Sample size estimates from simulation studies. Four studies estimated the sample size needed to detect a treatment effect using the ADAS-Cog 11 in a clinical trial of pre-dementia syndromes, and as summarized in Table 8 below, the ADAS-Cog 11 was never the outcome measure requiring the smallest sample size.⁶⁶⁻⁶⁹ A separate study found that increasing the proportion of APOEε4 allele carriers in clinical trial simulations, a method employed to try and increase the level of impairment of a sample, did not lead to meaningful increases in power to detect a treatment effect with the ADAS-Cog 11.³¹ Furthermore, the ADAS-Cog 11 failed to produce statistically significant treatment effects in several situations where one was hypothesized to be present based on other indicators of disease progression.³²

Table 8 Sample Size Estimates to Detect Treatment Effects in Pre-Dementia Clinical Trials

| First Author & Publication Year | Study Details | ADAS-Cog 11 rank versus other outcome measures |
|---------------------------------|--|--|
| Caroli 2015 ⁶⁶ | Estimate n per treatment arm needed to detect 20% reduction in disease progression over 24 months, with beta=0.20, and alpha=0.05. | |
| | MCI with Aβ | 6th (n=568) of 6; best=brain atrophy rate (n=46) |
| | MCI with Hippocampal Atrophy | 6th (n >1000) of 6; best=brain atrophy rate (n=77) |
| Grill 2013 ⁶⁷ | Estimate n per treatment arm required to detect 25% treatment effect in cognitive measures over 24 and 36 months with beta=0.20 and alpha=0.05. Assessed different sample enrichment strategies. | |

| | | |
|-----------------------|---|--|
| | NC with APOEε4 allele, 36 months | 6th (No decline) of 6; best=RAVLT (n=499) |
| | MCI with APOEε4 allele, 24 months | 5th (n=908) of 6; best=CDR-SB (n=329) |
| | NC with Aβ, 36 months | 6th (n=420495) of 6; best=RAVLT (n=1090) |
| | MCI with Aβ, 36 months | 3rd (n=639) of 6; best=CDR-SB (n=292) |
| | NC with Total CSF Tau, 36 months | 6th (no decline) of 6; best=RAVLT (n=817) |
| | MCI with Total CSF Tau, 24 months | 4th (n=537) of 6; best=CDR-SB (n=292) |
| | NC with CSF Tau phosphorylated at threonine 181, 36 months | 6th (n=2200678) of 6; best=RAVLT total score (n=559) |
| | NC with CSF Tau phosphorylated at threonine 181, 24 months | 3rd (n=714) of 6; best=CDR-SB (n=296) |
| | NC with CSF Total Tau/Aβ, 36 months | 6th (no decline) of 6; best=RAVLT (n=559) |
| | MCI with CSF Total Tau/Aβ, 24 months | 4th (n=676) of 6; best=CDR-SB (n=258) |
| | NC with CSF pTau/Aβ, 36 months | 6th (n=214455) of 6; best=RAVLT (n=552) |
| | MCI with CSF pTau/Aβ, 24 months | 3rd (n=696) of 6; best=CDR-SB (n=313) |
| | NC with Brain Glucose Hypometabolism, 36 months | 6th (n=13136) of 6; best=CDR-SB (n=1039) |
| | MCI with Brain Glucose Hypometabolism, 24 months | 3rd (n=357) of 6; best=MMSE (n=314) |
| | NC with Hippocampal Volume, 36 months | 6th (n=21359) of 6; best=CDR-SB (n=1057) |
| | MCI with Hippocampal Volume, 24 months | 5th (n=754) of six; best=CDR-SB (n=300) |
| | NC with Lateral Ventricle Volume, 36 months | Tied for 6th (no decline) of 6; best=RAVLT delayed recall (n=1039) |
| | MCI with Lateral Ventricle Volume, 24 months | 3rd (n=666) of 6; best=CDR-SB (n=381) |
| Ho 2010 ⁶⁸ | Estimate n per treatment arm to measure 25% reduction in rate of change over 12 months, beta=0.20 and alpha=0.05. | |
| | MCI | 4th (n=1183) of 5; best=Rate of Annual Brain Volume Loss (n=108) |

| | | |
|------------------------|--|--|
| Hua 2009 ⁶⁹ | Estimate n required to detect 25% reduction in rate of decline over 12 months with beta=0.20 or 0.10 and alpha=0.05. | |
| | MCI, 80% power | 6th (n=6797) of 6; best=Atrophy using symmetric Kullback-Leibler S9L5 distance (n=85) |
| | MCI, 90% power | 6th (n=9092) of 6; best=Atrophy using symmetric Kullback-Leibler S9L5 distance (n=114) |

Legend: See List of Abbreviations (pages xii to xvi).

3.2.8 Summary of ADAS-Cog 11 Performance in Pre-Dementia Studies

ADAS-Cog 11 scores in pre-dementia populations are driven primarily by the Word Recall and Word Recognition tasks, and age may influence scores for older adults with NC. Despite this, ADAS-Cog 11 scores do generally appear able to detect differences in cognitive ability in groups separated by an exposure that is expected to be associated with cognitive ability, although the magnitude of the differences detected tends to be small and are possibly attenuated by the nine tasks that demonstrated ceiling effects in pre-dementia populations. Responsiveness of the ADAS-Cog 11 to treatment effects appears low compared to other global outcome measures, and compared to outcome measures designed to assess subdomains of cognition or other aspects of dementia and pre-dementia syndromes. Nonetheless, caution must be maintained when interpreting these findings because an in-depth exploration of whether there truly should be an association between cognition or disease severity and any given exposure or treatment, and the potential magnitude and direction of these associations, was not explored. Overall, the ADAS-Cog 11 seems able to provide a measure of disease severity in pre-dementia syndromes, but there is room for improvement.

3.3 Modifications of the ADAS-Cog 11

This section reviews all modifications that have been made to the ADAS-Cog 11, as well as other outcome measures which have been combined with some or all of the individual

ADAS-Cog 11 tasks for the purpose of measuring disease severity in studies of dementia or pre-dementia syndromes. The organizational structure of this section is to introduce an outcome measure, review available information about responsiveness to baseline discrimination, disease progression, and treatment effects, and then summarize the performance of that outcome measure in comparison with the ADAS-Cog 11 (See Section 2.2.7 for responsiveness definitions). A visual representation of the modification history of the ADAS-Cog 11 is presented below in Figure 6, and a summary of each measure including coefficient of variation (CV) calculations can be found in Appendix A.

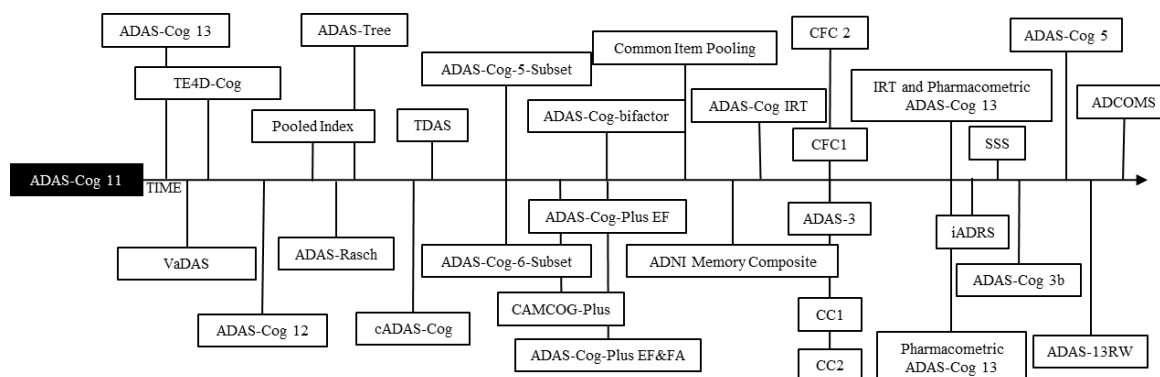


Figure 6 Timeline of ADAS-Cog 11 modifications.

3.3.1 ADAS-Cog 13

Mohs et al. (1997) identified several cognitive domains hypothesized to be important to detect treatment effects in clinical trials of antidementia drugs that are not assessed by the ADAS-Cog 11.⁷⁰ Accordingly, tests of attention and concentration, planning and executive function, verbal memory, nonverbal memory, and praxis were considered for addition to the ADAS-Cog 11.⁷⁰ Recommendations about which specific tests to add to the ADAS-Cog 11 were based on assessments of reliability, influence of age and education on change scores, learning effects (one month interval), ability to assess full range of dementia severity, floor and ceiling effects, and ability to measure 12 month longitudinal change in 64 subjects with NC, 50 subjects with mild AD, 47 subjects with moderate AD, and 46 subjects with moderately severe AD.⁷⁰ In summary, the authors

recommended adding items such as a four-trial learning plus delayed word recall task, one or two simple mazes, or number cancellation tasks for clinical trials involving subjects with mild AD or pre-dementia.⁷⁰ A commonly used adaptation of these suggestions is the ADAS-Cog 13 which includes all ADAS-Cog 11 items as well as a test of delayed word recall and a number cancellation or maze task. Errors on the additional task are summed together with scores from the original 11 tasks to give a final ADAS-Cog 13 score from 0 to 85.

Baseline discrimination. Skinner et al. (2012) found the mean score on the ADAS-Cog 13 was lower for 229 subjects with NC (mean=9.5, SD=4.1) than 394 subjects with MCI (mean=18.6, SD=6.2) and 187 subjects with AD (mean=28.9, SD=7.6).³ In a separate analysis, Podhorna et al. (2016) found the mean ADAS-Cog 13 score for 382 subjects with MCI (mean=15.23, SD=6.68) was lower than that of 97 subjects with mild AD (mean=29.91, SD=7.44).¹² Podhorna et al. (2016) further divided the 382 subjects with MCI into two groups depending on whether a CSF or APOEε4 allele biomarker of AD pathology was present.¹² The 206 subjects with MCI and an indication of AD pathology (enriched subgroup) had worse scores at baseline (mean=17.52, SD=6.81) on the ADAS-Cog 13 than the 176 subjects with MCI but no such AD biomarkers present (mean=12.55, SD=5.43).¹² Statistical significance of the above differences was not tested.

Disease progression. Hobart et al. (2009) used Rasch Analysis to compare the ADAS-Cog 11 and ADAS-Cog 13 on scale performance and person measurements in 371 subjects with MCI and 217 subjects with AD.⁷¹ Although they found that the ADAS-Cog 13 evaluates more cognitive domains than the ADAS-Cog 11, it was not better at measuring clinically significant changes in subjects with MCI.⁷¹ Podhorna et al. (2016) had similar results. They found little change on the ADAS-Cog 13 for 382 subjects with MCI over 24 months (mean change=1.34 points) or for 168 subjects with MCI over 36 months (mean change=2.59 points).¹² There was slightly more change detected in the enriched MCI subgroup (mean 24 month change=2.63 points; mean 36 month change=5.02 points), and no meaningful change on the ADAS-Cog 13 in the non-enriched MCI subgroup (mean 24 month change=-0.18 points, mean 36 month change=-0.15 points).¹² Among 97 subjects with mild AD there was a modest change in mean

ADAS-Cog 13 score over 12 months (mean change=4.35 points) and among 38 subjects with AD over 24 months (mean change=9.46 points).¹² The SRM for change over 24 months in 382 subjects with MCI, adjusting for baseline age, baseline MMSE score, sex, and APOEε4 allele, was 0.39 (95% CI 0.16, 0.60) for the ADAS-Cog 13 compared to 0.37 (95% CI 0.15, 0.57) for the ADAS-Cog 11.¹² The SRM for change over 12 months in 97 subjects with AD was 0.98 (95% CI 0.58, 1.26) for the ADAS-Cog 13 and 0.87 (95% CI 0.46, 1.13) for the ADAS-Cog 11.¹² Skinner et al. (2012) found the Z-statistic for change over time in 394 subjects with MCI was slightly larger for the ADAS-Cog 13 (Z=10.70) than for the ADAS-Cog 11 (Z=9.44), adjusting for age, education, gender, and APOEε4 allele.³ Raghavan et al. (2013) also found the ADAS-Cog 13 had larger standardized two-year change than the ADAS-Cog 11 in an MCI sample.¹⁹

Treatment effect. Skinner et al. (2012) found that the estimated sample size per group to detect a 25% decrease over 12 months in subjects with MCI with 80% power and an alpha of 0.05 was smaller for the ADAS-Cog 13 (n=900) than for the ADAS-Cog 11 (n=1230).³ Raghavan et al. (2013) found the estimated sample size to detect a hypothetical 25% treatment effect over 2 years in subjects with MCI with 80% power was also smaller for the ADAS-Cog 13 (n=582) than for the ADAS-Cog 11 (n=772).¹⁹

Summary. The ADAS-Cog 13 appeared able to discriminate between groups of subjects with MCI and mild AD at one point in time. For subjects with AD, the responsiveness of the ADAS-Cog 13 to disease progression was better than that of the ADAS-Cog 11. For subjects with pre-dementia syndromes, the ADAS-Cog 13 demonstrated similar or only slightly better responsiveness to disease progression than the ADAS-Cog 11. Responsiveness to treatment effects in MCI was better for the ADAS-Cog 13 than the ADAS-Cog 11.

3.3.2 Vascular Dementia Assessment Scale

To address the need for a primary outcome measure for clinical trials in Vascular Dementia (VaD), Ferris et al. (1999) suggested using the ADAS-Cog 11 as a starting point because many of the cognitive domains affected by VaD are also affected in AD.⁷² Cognitive domains important for VaD include memory, attention, processing speed,

visuospatial function, language, executive function, and abstraction.⁷² The original VaDAS included items suggested by Mohs et al. (1997) for the ADAS-Cog 13.^{70,72} An updated version had additional items to further target frontal lobe functions.⁷³ This updated VaDAS includes all tasks from the ADAS-Cog 11 as well as a delayed recall portion added to the Word Recall task (memory), two number cancellation tasks (attention), a maze (executive function), symbol digit modalities (attention/concentration), backwards digit span (working memory), and animal category retrieval (verbal fluency) tasks.⁷³ At the time of development, evaluation of the VaDAS was left to be done in future clinical trials.⁷³

Disease Progression. The VaDAS showed improvement over 18 weeks for both placebo and donepezil groups in an 18 week randomized clinical trial of donepezil versus placebo for 168 subjects with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (least squares mean change from baseline: placebo=-0.81, donepezil group=-0.85).⁷⁴

Treatment effect. No statistically significant treatment effect was found using the VaDAS, which was congruent with a version of the ADAS-Cog 13 (ADAS-Cog 11, number cancellation, maze), ADAS-Cog 11, and MMSE.⁷⁴

Summary. Theoretically the VaDAS should be better than the ADAS-Cog 11 at detecting VaD dysfunction and assessing change in this dysfunction over time; however, further studies are needed to definitively evaluate whether this is true. Analyses of baseline discrimination have also not been performed.

3.3.3 ADAS-Cog 12

A common modification to the ADAS-Cog 11 is to add a Delayed Word Recall task which provides a subject three trials to recall as many of the ten words from the Word Recall task after a period of time (delay).¹⁷ The task is scored from 0 to 10 based on the number of words not recalled (errors), and added to the ADAS-Cog 11 score to give a total score of 0 to 80.

Lowe et al. (2015) used IRT methods with 788 subjects ranging from pre-dementia syndromes (Subjective Cognitive Impairment (SCI), MCI) to severe stages of AD to determine where along the continuum of disease progression the Delayed Word Recall task provides the most information about cognitive ability, and if it alters the measurement precision of the ADAS-Cog 11.⁷⁵ ICCs from the IRT analysis showed that in general memory tasks of the ADAS-Cog 12 are the most sensitive to the earlier stages of disease progression, and the Delayed Word Recall task provides the most information in the mildest range of cognitive impairment.⁷⁵ Area under the curve (AUC) analyses found statistically significant differences in the overall average distance between the ICC for the Delayed Word Recall task and for the Word Recall task.⁷⁵ The Delayed Word Recall task does not however have much sensitivity for more severe cognitive dysfunction such as that seen with AD.⁷⁵ Floor effects (10 errors) on the Delayed Word Recall task were seen for 9% of the MCI group and 52% of the AD group at baseline.⁷⁵

Baseline discrimination. Grundman et al. (2004) found statistically significant different mean scores in the ADAS-Cog 12 between groups of subjects with aMCI (n=769), NC (n=107), very mild AD (n=122), and mild AD (n=183). Furthermore, subjects with MCI performed an average 2.1 SD higher than subjects with NC on the Delayed Word Recall task compared to an average of 1.8 SD higher on the original immediate Word Recall task.¹⁷ Sano et al. (2011) showed that 111 subjects with AD had significantly higher mean scores (two samples *t*-test, $t = 15.3$, $P < 0.001$) on the ADAS-Cog 12 (mean=33.27 points, SD=10.3) than 259 subjects with MCI (mean=17.22 points, SD=5.9).⁷ Test information curves for the ADAS-Cog 11 and ADAS-Cog 12 show that both scales are maximally precise around mild to moderate AD, but the ADAS-Cog 12 is more precise, or holds more information about underlying cognitive impairment, in earlier stages of disease progression than the ADAS-Cog 11.⁷⁵ The ADAS-Cog 12 maintains similar precision to the ADAS-Cog 11 for more severe stages of cognitive impairment, namely AD.⁷⁵ Labos et al. (2011) compared performance on the MMSE, ADAS-Cog 11, traditional memory ADAS-Cog 11 tasks (immediate Word Recall and Word Recognition), and an additional Delayed Word Recall task in 230 subjects divided into NC, SCI, amnesic MCI (aMCI), multidomain MCI, and dementia categories.⁷⁶ The scores from all four tests or subtasks were able to distinguish the group of subjects with dementia from subjects with NC, SCI,

aMCI, and multidomain MCI.⁷⁶ The MMSE did not significantly discriminate between any other groups. Scores on the ADAS-Cog 11 as well as the isolated memory tasks were similar for groups with NC and SCI, but NC group scores were significantly better than the two MCI subtype groups, which were not distinguishable.⁷⁶ The Delayed Word Recall task score was comparable for groups with NC and SCI, but the group with NC scored significantly worse than both MCI subtype groups.⁷⁶ Furthermore, the aMCI scores were significantly worse than those for multidomain MCI.⁷⁶

Disease progression. Sano et al. (2011) found that 12 month unadjusted change scores were significantly different between MCI and AD groups for the ADAS-Cog 11 ($t=4.26$, $P<0.001$) and ADAS-Cog 12 ($t=3.89$, $P<0.001$), but the Delayed Word Recall task on its own was not ($t=-0.45$, $P=0.654$).⁷ Among the MCI group, the 12 month SRM was lower for the ADAS-Cog 11 (0.142) than for the ADAS-Cog 12 (0.160).⁷ The ratio of the SRM for the ADAS-Cog 12 divided by the SRM for the ADAS-Cog 11 was used to show that including Delayed Word Recall with the ADAS-Cog 11 increased the SRM by 12% (more responsive).⁷ For the AD group, the 12 month SRM was similar between the ADAS-Cog 11 (0.589) and ADAS-Cog 12 (0.569).⁷

Treatment effect. The estimated sample size required to detect a 33% treatment effect in MCI with 80% power was over 600 subjects lower for the ADAS-Cog 12 than the ADAS-Cog 11.⁷ In contrast, the ADAS-Cog 12 did not outperform the ADAS-Cog 11 for estimations of sample size needed for a trial of AD.⁷

Summary. The ADAS-Cog 12 has demonstrated the ability to discriminate between groups of subjects with MCI and AD, as well as between MCI subtypes. The ADAS-Cog 12 demonstrated more responsiveness to disease progression and treatment effects in MCI than the ADAS-Cog 11. Further along the disease continuum responsiveness to disease progression and treatment effects of the ADAS-Cog 11 and 12 were comparable.

3.3.4 Test for the Early Detection of Dementia from Depression – Cognitive

The Test for the Early Detection of Dementia from Depression (TE4D) was initially developed in the German language to differentiate early dementia from depression.⁶ Mahoney et al. (2005) modified the TE4D with the intention of using it as a screening tool to detect MCI in English-speaking populations with AD (TE4D-Cog).⁶ The TE4D-Cog is scored from 0 to 45, and has eight items among seven subscales which assess immediate recall, semantic memory, clock drawing, category fluency, orientation, and following commands (from ADAS-Cog 11).⁶ The TE4D-Cog was tested in a sample of 178 subjects with AD and 25 subjects with NC, where it was found to have good concurrent criterion validity with the ADAS-Cog 11 ($r=-0.90$, $P<0.001$) and MMSE ($r=0.92$, $P<0.001$), high inter-rater reliability, and good internal consistency.⁶

Baseline discrimination. Twenty-five subjects with NC scored significantly better on the TE4D-Cog than 178 subjects with AD both in terms of overall score (Mann-Whitney U test (U)=24.0, $P<0.001$), and each of the seven subscales ($P<0.001$).⁶ The ability of the TE4D-Cog to serve as a screening tool for dementia was compared with the MMSE by assessing sensitivity and specificity at different cut-points and calculating the area under the receiver operating characteristic (ROC) curve.⁶ The cut-point giving maximum (100%) sensitivity for the TE4D-Cog (score ≥ 35) corresponded to a specificity of 84.0%. Using a cut-point for maximum sensitivity of the MMSE (score ≥ 29) corresponded to a specificity of 32.0%.⁶ When set at maximum specificity (100%), the sensitivity of the TE4D-Cog drops to 79.5% and the MMSE to 65.9%. The AUC for the TE4D-Cog was 0.98, and 0.96 for the MMSE.⁶

Disease progression. In a subsample of 148 subjects with AD, baseline (mean=16.2, SD=11.1) and six month follow-up scores (mean=14.2, SD=10.8) on the TE4D-Cog were correlated ($r=0.90$, $P<0.001$) and there was a statistically significant worsening in scores over time (Wilcoxon signed rank test; $Z=-4.9$, $P<0.001$).⁶

Summary. The TE4D-Cog demonstrated the ability to discriminate between NC and AD groups, and demonstrated responsiveness to disease progression in AD. Responsiveness to treatment effects and comparison with the ADAS-Cog 11 was not tested.

3.3.5 Pooled Index

Carusone et al. (2006) were the first to add measures of functionality rather than just measures of cognition to the ADAS-Cog 11. They used data from a clinical trial involving 101 subjects with mild to moderate AD, and combined the following six scales using a pooled index approach: ADAS-Cog 11, Geriatric Depression Scale, Dysfunctional Behaviour Rating Instrument (DBRI), MMSE, Activities of Daily Living (ADL), and DBRI frequency.⁷⁷

Treatment effect. Effect sizes were calculated for each individual subscale measure as well as the Pooled Index for 3, 6, and 12 months of follow-up (Effect size=linear regression coefficient/SE of linear regression coefficient). None of the individual subscale measures demonstrated a statistically significant treatment effect at more than one time point.⁷⁷ The Pooled Index found a statistically significant treatment effect at the 3 month and 12 month, but not 6 month, follow-up assessments.⁷⁷

AUC analyses of individual scores plotted against time were performed for both the standardized ADAS-Cog 11 and the Pooled Index. The standardized ADAS-Cog 11 showed a statistically significant difference between treatment and placebo groups at the finite time period of 6 months, but not when assessing the 12 month time period as a whole.⁷⁷ The Pooled Index showed statistically significant difference between placebo and treatment groups over the entire 12 month period, and at the individual time points of 3 and 12 months, but not at 6 months.⁷⁷

Summary. The Pooled Index was more responsive to treatment effects than the ADAS-Cog 11 in a clinical trial for mild to moderate AD. Responsiveness to baseline discrimination and to disease progression was not explored.

3.3.6 ADAS-Rasch

Wouters et al. (2008) identified three problematic aspects of the ADAS-Cog 11 scoring methodology, whereby the total score is arrived at by summing points across tasks without recognition of how these individual tasks or subtask item response options may differ.⁷⁸ Specifically, they found ADAS-Cog 11 tasks do not have equal measurement precision, several subtask item response categories are disordered in terms of difficulty, and a difference of a certain number of points at the low end of the scoring range does not equal the same amount of difference in cognitive ability as a difference of the same number of points at the higher end of the scoring range.⁷⁸ This last limitation suggests that summed ADAS-Cog scores should not be treated as an interval-ratio level measure and analyzed using parametric statistics such as t-tests and linear regression models.

To address the first two limitations Wouters et al. (2008) developed an alternate scoring method for the ADAS-Cog 11 using Rasch analysis. In brief, response categories with the same level of difficulty for each task on the ADAS-Cog 11 were collapsed so that the ADAS-Rasch has hierarchically ordered categories, and each task is weighted according to its measurement precision.⁷⁸ The total possible score for each task of the ADAS-Rasch is the product of the number of different categories of difficulty present for the items of that task and the weight assigned to the task: Word Recall (total possible points for ADAS-Cog 11=10 versus total possible points for ADAS-Rasch=12), Naming (5 versus 6), Commands (5 versus 8), Constructional Praxis (5 versus 4), Ideational Praxis (5 versus 6), Orientation (8 versus 6), Word Recognition (12 versus 3), Remembering Test Instructions (5 versus 5), Language Ability (5 versus 5), Word-Finding Difficulty (5 versus 4), Comprehension (5 versus 5), total score (70 versus 64).⁷⁸ ADAS-Rasch scores are backwards-compatible to a Classical Test Theory-derived ADAS-Cog 11 sum score. The third scoring limitation (non-equal intervals) remains for both the ADAS-Cog 11 and the ADAS-Rasch.

The ADAS-Rasch was developed from baseline data of the placebo arms of three clinical trials that included 706 subjects with mild to moderate dementia.⁷⁸ External criterion validation was performed in 456 patients from a different trial with similar inclusion and

exclusion criteria.⁷⁸ There was a high correlation between ADAS-Rasch and ADAS-Cog 11 scores ($r=0.93$), and a moderate correlation between ADAS-Rasch and MMSE scores ($r=-0.72$).⁷⁸

Baseline discrimination. One-to-one correspondence between the ADAS-Rasch total score and level of cognitive impairment was demonstrated in both the development and external-validation samples.⁷⁸ This one-to-one correspondence was not present for the ADAS-Cog 11, meaning two individuals with the same score may have different levels of cognitive impairment.⁷⁸ In a later study, Skinner et al. (2012) found that mean scores on the ADAS-Rasch were lower for 229 subjects with NC (mean=4.8, SD=3.5) than for 394 subjects with MCI (mean=11.8, SD=5.5), or 187 subjects with AD (mean=19.5, SD=7.4).³ Crane et al. (2012) found comparable results.²⁵ Statistical tests of these differences were not performed.

Disease progression. Skinner et al. (2012) found the Z-score for change over time in 394 subjects with MCI, adjusted for age, education, gender, and APOEε4 allele, was smaller for the ADAS-Rasch ($Z=8.50$) than the ADAS-Cog 11 ($Z=9.44$).³ Similar analyses performed by Crane et al. (2012) found adjusted Z-scores for time were smaller for the ADAS-Rasch than ADAS-Cog 11 in NC (ADAS-Rasch=3.10, ADAS-Cog 11=3.20), MCI (ADAS-Rasch=-10.51, ADAS-Cog 11=-10.78), and AD (ADAS-Rasch=-11.28, ADAS-Cog 11=-12.25) samples.²⁵

Treatment effect. Skinner et al. (2012) found that the estimated sample size per group to detect a 25% decrease over 12 months in MCI with 80% power and an alpha of 0.05 was larger for the ADAS-Rasch ($n=1409$) than for the ADAS-Cog 11 ($n=1230$).³ Crane et al. (2012) found that the ADAS-Rasch required a larger estimated sample size than the ADAS-Cog 11 to detect a 25% decrease over 12 months, with 80% power and an alpha of 0.05, for NC (41,295 versus 37,971), MCI (1692 versus 1651), and AD (346 versus 242).²⁵

Summary. The ADAS-Rasch improved two of three problem areas with traditional ADAS-Cog 11 scoring methodology, and appeared to demonstrate better baseline discrimination than the ADAS-Cog 11. Responsiveness to disease progression and

treatment effects was worse for the ADAS-Rasch than the ADAS-Cog 11 in NC, MCI, and AD.

3.3.7 ADAS-Tree

Llano et al. (2011) developed an alternative weighting scheme for scoring the ADAS-Cog 13 to identify subjects with MCI who have a high risk of converting to AD.¹⁸ The rationale for this was to increase the efficiency of a clinical trial by using conversion from MCI to AD as an outcome, and then enrolling subjects with a particularly high risk of this conversion.¹⁸ A second purpose of the ADAS-Tree is to discriminate between subjects with different levels of disease severity at the start of a clinical trial.¹⁸ Results for this baseline discriminative ability will be reviewed, but not results pertaining to the ability of the ADAS-Tree to predict conversion from MCI to AD as risk prediction is less relevant for this thesis.

To develop the ADAS-Tree, the Random Forests (RF) tree-based algorithm was used to derive weights for each task of the ADAS-Cog 13 based on their ability to discriminate between subjects with NC, MCI, and AD. Briefly, ten thousand bootstrap datasets were taken from baseline data of 229 subjects with NC, 397 subjects with MCI, and 193 subjects with AD.¹⁸ The RF algorithm was applied in each bootstrap dataset to develop a classification tree for NC, MCI, and AD diagnostic categories.¹⁸ Each bootstrap dataset was the same size as the original sample, but because datasets were obtained using random sampling with replacement, about one third of the original sample was not selected for any given bootstrap (some observations were sampled multiple times).¹⁸ These left out samples were used to obtain an estimate of predictive accuracy by comparing diagnoses predicted by the majority of classification trees (RF model) with original diagnoses.¹⁸ Weights for each task of the ADAS-Cog 13 were derived by comparing the predictive accuracy of the RF model fit using the full ADAS-Cog 13 to the predictive accuracy of a RF model fit when one ADAS-Cog 13 task was replaced by noise, repeated for all tasks.¹⁸ Tasks that led to a large decrease in predictive accuracy when excluded were given the highest weights in the ADAS-Tree as this reflects a

relatively large contribution to the ability of the ADAS-Cog 13 to discriminate between NC, MCI, and AD.¹⁸

Item weights of the ADAS-Tree are: 1.05 Word Recall, 0.38 Commands, 0 Construction, 1.17 Delayed Word Recall, 0.61 Naming, 0.13 Ideational Praxis, 1.13 Orientation, 0.41 Word Recognition, 0.54 Recall Instructions, 0.49 Spoken Language, 0.69 Word Finding, 0.39 Comprehension, 0.69 Number Cancellation.¹⁸

Baseline discrimination. The ADAS-Tree was able to discriminate between NC, MCI, and AD diagnostic categories ($P < 0.0001$).¹⁸ Furthermore, the Kruskal-Wallis test statistic used to assess the magnitude of difference between these categories was larger for the ADAS-Tree (401.1) than the ADAS-Cog 13 (393.3), ADAS-Cog 11 (378.9), and MMSE (368.8).¹⁸ A separate study found ADAS-Tree scores were lower for 229 subjects with NC (mean=7.9 points, SD=3.5) than 394 subjects with MCI (mean=15.9 points, SD=5.1), and 187 subjects with AD (mean=24.2 points, SD=5.6).³ A third study similar differences in scores between NC, MCI, and AD diagnostic categories.²⁵

Disease progression. Skinner et al. (2012) found that for 394 subjects with MCI the ADAS-Tree had a larger Z-score for time ($Z=12.04$) than the ADAS-Cog 11 ($Z=9.44$), adjusted for age, education, gender, and APOEε4 allele.³ Crane et al. (2012) also found that the ADAS-Tree had a larger adjusted Z-score for time than the ADAS-Cog 11 in MCI (ADAS-Tree: $Z=-13.67$, ADAS-Cog 11: $Z=-10.78$) and AD (ADAS-Tree: $Z=-14.05$, ADAS-Cog 11: $Z=-12.25$), but not NC (ADAS-Tree: $Z=0.73$, ADAS-Cog 11: $Z=3.20$) samples.²⁵

Treatment effect. Skinner et al. (2012) found that the estimated sample size per group to detect a 25% decrease over 12 months in MCI with 80% power and an alpha of 0.05 was smaller for the ADAS-Tree ($n=733$) than for the ADAS-Cog 11 ($n=1230$).³ Crane et al. (2012) demonstrated that the ADAS-Tree required a larger estimated sample size than the ADAS-Cog 11 to detect a 25% decrease over 12 months, with 80% power and an alpha of 0.05, for subjects with NC (573,996 versus 37,971), and a smaller estimated sample size than the ADAS-Cog 11 for subjects with MCI (981 versus 1651) or AD (214 versus 242) hypothetical clinical trials.²⁵

Summary. The ADAS-Tree demonstrated greater baseline discrimination ability than the ADAS-Cog 11 for detecting a difference among NC, MCI, and AD diagnostic categories. Responsiveness to disease progression and treatment effects appears to be improved by the ADAS-Tree in MCI and AD, but not NC.

3.3.8 Computerized ADAS-Cog

The National Institute on Aging funded the development of a computerized version of the ADAS-Cog (cADAS-Cog) to try to increase consistency between and decrease errors made by administrators of the ADAS-Cog.⁷⁹ The cADAS-Cog includes a computerized version of all ADAS-Cog 11 items plus Delayed Recall, Number Cancellation, and Maze tasks. It is administered using a Computerized Multiphasic Interactive Neurocognitive Dual Display System (CMINDS).⁷⁹ The first step at any testing session is for the subject to practice using CMINDS via a Perception Response Evaluation (PRE) module.⁷⁹ A secondary purpose of the PRE module is to ensure subjects have sufficient perceptual and response abilities to take the computerized test.⁷⁹ Next, the cADAS-Cog is administered on one monitor display while the test administrator uses the second monitor to control the speed of the testing, request repeated test instructions, and receive information on the subject's progress throughout the test.⁷⁹

A sample of 88 subjects with mild to moderate AD were administered both the computerized and paper ADAS-Cog versions three times, four months apart.⁷⁹ Different versions were given on alternate time points, each one month apart.⁷⁹ Both computerized and paper tests took approximately 44 minutes to administer.⁷⁹ High concurrent criterion validity between the cADAS-Cog and paper version total scores, and all individual task scores was suggested by ICCs (all $P < 0.001$), Pearson's correlation coefficients (all $P < 0.01$), and paired sample t-tests of differences between intra-subject scores (all $P > 0.10$).⁷⁹ High test-retest reliability was found over approximately five month ($P < 0.001$) and ten month periods ($P < 0.001$).⁷⁹ Paired sample t-tests showed that the reliability across cADAS-Cog scores was significantly better than that of the paper administration method over five and ten month periods (5 month: mean cADAS-Cog ICC=0.87, mean paper ICC=0.80, $t=2.88$, $P < 0.02$; 10 month: mean cADAS-Cog ICC=0.83, mean paper

ICC=0.77, $t=2.54$, $P<0.03$).⁷⁹ Agreement was also demonstrated with a Bland-Altman plot of the differences between total scores.⁷⁹

Summary. The cADAS-Cog improved the reliability and standardization of an extended version of the ADAS-Cog 13, and may be considered as an alternative mode of administration especially when ADAS-Cog scores from multiple different administrators are going to be compared. Explicit tests of responsiveness were not conducted.

3.3.9 Touch Panel-Type Dementia Assessment Scale

Inoue et al. (2011) created a computerized version of a modified ADAS-Cog 11 that can be administered in 30 minutes.⁸⁰ This Touch Panel-Type Dementia Assessment Scale (TDAS) has a 14" touch panel display and includes the seven tasks of the ADAS-Cog 11 which they could computerize (Word Recognition, Following Commands, Orientation, visual-spatial perception (modified Constructional Praxis), Naming Fingers, object recognition (modified Naming Objects), accuracy of the order of a process (modified Ideational Praxis) as well as tests for money calculation and clock time recognition (non-digital).⁸⁰ The scoring range is 0 to 101, with lower scores indicating worse performance. A limitation of the TDAS is that people with severe AD or visual and/or hearing impairment require assistance or may not be able to finish the test.⁸⁰

Thirty-four subjects with AD were administered both the TDAS and a paper version for concurrent criterion validation analyses.⁸⁰ Total scores from the two tests were significantly correlated ($r=0.69$, $P<0.01$).⁸⁰ Kendall coefficients of concordance were calculated to assess agreement between six of the TDAS tasks and six of the paper ADAS-Cog 11 tasks.⁸⁰ Three tasks showed acceptable concordance [Word Recognition (0.57), Orientation (0.41), and Naming Objects and Fingers (0.32)], while three showed poor concordance [Following Commands, Constructional Praxis, Ideational Praxis (all Kendall's coefficients <0.3)].⁸⁰

Summary. The TDAS is a computerized test of cognitive ability which includes some modified items of the ADAS-Cog 11. Preliminary tests of agreement were mixed. Further tests would help to establish responsiveness of the TDAS.

3.3.10 Computerized Adaptive Testing of the Cambridge Cognitive Examination – Plus

The Cambridge Cognition Examination (CAMCOG)-Plus is composed of a battery of neuropsychological tests including the ADAS-Cog 11.⁸¹ Wouters et al. (2011) used Computerized Adaptive Testing (CAT) to administer the CAMCOG and CAMCOG-Plus to 41 subjects with NC, 21 subjects with MCI, and 22 subjects with dementia to see whether the CAT version maintains diagnostic accuracy while decreasing length of administration.⁸¹

The CAT procedure begins by asking subjects a series of standard questions which an internal algorithm uses to estimate cognitive ability.⁸¹ Each time a question is answered throughout the entirety of the testing procedure (i.e. during CAMCOG or CAMCOG-Plus administration) the algorithm updates the estimate of cognitive ability and uses the response to select the difficulty of the next question to be administered.⁸¹ Correct responses lead to the administration of more difficult questions while incorrect responses lead to the administration of easier ones.⁸¹ Difficulty levels of test items were initially estimated using a one parameter logistic model.⁸¹ The updating process is continued until 25 items are administered or a standard error of measurement corresponding to 90% reliability for cognitive ability is reached.⁸¹

In the original sample an estimate of cognitive ability was reached using the CAT CAMCOG-Plus after administering 53% fewer items than are included in the full test battery.⁸¹ Time to administer was reduced by 54%.⁸¹ The CAT CAMCOG-Plus had excellent agreement for estimating cognitive ability with the paper CAMCOG-Plus (ICC 0.98, $P < 0.001$) and the paper CAMCOG (ICC 0.99, $P < 0.001$).⁸¹ Concurrent criterion validity was found between the CAMCOG-Plus and MMSE (Spearman's $\rho = 0.80$, $P < 0.001$) and Informant Questionnaire on Cognitive Decline in the Elderly (Spearman's $\rho = -0.54$, $P < 0.025$).⁸¹

Baseline Discrimination. The CAT CAMCOG-Plus but not CAT CAMCOG was better at discriminating between the diagnostic categories of NC, MCI, and dementia than the MMSE, as assessed by AUCs and optimal sensitivity and specificity values.⁸¹

Summary. The CAT CAMCOG-Plus demonstrated the ability to discriminate between NC, MCI, and dementia diagnostic categories. Direct comparisons with the ADAS-Cog 11 were not performed, nor were assessments of responsiveness to disease progression and treatment effects.

3.3.11 ADAS-Cog-5-Subset and ADAS-Cog-6-Subset

Ihl et al. (2012) used a subsetting analysis approach to develop two separate subsets of ADAS-Cog 11 tasks based on the ability of individual tasks to detect a treatment effect in three 24 week randomized controlled trials of a total of 855 subjects with mild to moderate AD.⁸² The objective of the subsetting analysis was to remove tasks from the ADAS-Cog 11 that demonstrated low sensitivity for detecting a treatment response.⁸²

The first step of the subsetting analysis was to calculate the pre-post difference for all ADAS-Cog 11 tasks.⁸² If the pre-post difference on the task score was less than or equal to 0 (did not get worse over time; treatment responder) a binary variable for ‘response’ was given the value 1.⁸² If the pre-post difference was greater than 0 (subject got worse over the course of the study; treatment non-responder) the binary ‘response’ variable was coded as 0.⁸² Importantly, a “responder” was defined as a subject who showed no worsening on any task of a given subset of tasks over the course of the study.⁸²

Responders could be in the placebo or treatment group. A mathematical algorithm was then used to identify subsets of ADAS-Cog 11 tasks which could identify groups of responders, and Wilcoxon rank-sum tests were used to find subsets with statistically significant differences in the proportion of responders between treatment and placebo groups.⁸² The subset with the lowest *P*-value was selected as the collection of ADAS-Cog 11 tasks with the most potential for detecting a treatment response.⁸²

The primary result of this analysis was the ADAS-Cog-5-Subset: Ideational Praxis, Remembering Test Instructions, Language, Comprehension, and Word Finding Difficulty tasks. Internal consistency of the ADAS-Cog-5-Subset (Chronbach’s alpha=0.81) was close to that of the ADAS-Cog 11 (Chronbach’s alpha=0.82).⁸²

Tasks not selected for the primary subset were combined to create the ADAS-Cog 6-Subset: Word Recall, Naming Objects and Fingers, Commands, Constructional Praxis, Orientation, and Word Recognition.⁸²

Treatment effect. The ADAS-Cog-5-Subset found statistically significant differences in the proportion of responders in the treatment compared to control groups for the overall study population ($P=0.0001$), as well as subgroups of subjects with mild AD ($P=0.01$), and moderate AD ($P=0.01$).⁸² The ADAS-Cog-6-Subset found statistically significant treatment effects for the overall study population ($P=0.0016$) and the moderate AD subgroup ($P=0.0002$), but not among the subgroup of mild AD subjects ($P=0.53$).⁸² The ADAS-Cog 11 found no statistically significant difference between the proportion of responders in the treatment versus control group.⁸²

Summary. Both the ADAS-Cog-5-Subset and the ADAS-Cog-6-Subset were more responsive than the ADAS-Cog 11 to treatment effects in AD. The ADAS-Cog-5-Subset was the most sensitive for detecting a memantine treatment response in mild AD. The ADAS-Cog-6-Subset was the most sensitive for detecting a memantine treatment response in moderate AD. Although subsetting analysis requires longitudinal data from a clinical trial, future studies may evaluate the ability of the Subsets to detect within-person change over time in observational studies. One limitation, common to other validation analyses, is that the Subsets were statistically optimized for sampling and measurement error of the test dataset and will not likely have the same performance characteristics in a new study.

3.3.12 The ADAS-Cog-Plus (ADAS-Bifactor, ADAS-Plus-EF, ADAS-Plus-EF&FA)

Skinner et al. (2012) used two strategies to modify the ADAS-Cog 13 to try and improve responsiveness to changes in MCI.³ First, alternative weights to tasks of the ADAS-Cog 13 were applied based on latent trait analysis with IRT. This resulted in a bi-factor model that accounted for correlations between Word Recognition and Word Recall tasks, and for correlations between the four examiner-rated tasks.³ The variance of the primary factor was fixed at one, and loadings were freely estimated.³ Scores for follow-up visits were

computed using item parameters from this baseline model transformed to a standard normal distribution (mean=0 and SD=1).³ Second, two other variants of the ADAS-Cog 13 were created by adding tasks to assess Executive Functioning (EF) and informant reports of daily function (FA). The ADAS-Plus-EF consists of the ADAS-Cog 13 plus an additional task for category (vegetable) fluency.³ The ADAS-Plus-EF&FA consists of the ADAS-Cog 13 plus category (vegetable) fluency, Trail Making Tests (TMT) A and B, Digit Symbol Substitution (DSS) Test, and five Pfeffer FAQ items.³ These modifications were developed using data from 811 subjects with a range of cognitive abilities, validated in a subset of 394 subjects with MCI, and then compared to the ADAS-Cog 11, ADAS-Cog 13, ADAS-Rasch (Section 3.3.6), and ADAS-Tree (Section 3.3.7).³

Baseline discrimination. Plots of test information curves (Monte Carlo integrated test information versus cognitive ability) showed that the ADAS-Plus-EF&FA model had the highest test information over all levels of cognitive ability, followed by the ADAS-Plus-EF, and then the ADAS-Bifactor.³ In general, the amount of information any of the three variants held about cognitive ability increased as cognitive ability worsened.³

Disease progression. The Z-score for change over time in the validation sample adjusting for age, education, gender and APOEε4 allele was larger for the ADAS-Plus-EF&FA (Z=11.81) than the ADAS-Plus-EF (Z=10.61), ADAS-Bifactor (Z=10.26), and ADAS-Cog 11 (Z=9.44).³

Treatment effect. Estimated sample sizes to detect a 25% change in cognition over 12 months with 80% power and alpha of 0.05 were calculated.³ The ADAS-Plus-EF&FA required a smaller sample size (n=547) than the ADAS-Plus EF (n=883), ADAS-Bifactor (n=1103), and ADAS-Cog 11 (n=1230).³

Summary. The ADAS-bifactor, ADAS-Plus-EF, and ADAS-Plus EF&FA all demonstrated the ability to provide information about cognitive ability across various levels of cognitive impairment, suggesting they may be responsive to baseline discrimination. The ADAS-Bifactor, ADAS-Plus-EF, and ADAS-Plus EF&FA all showed superior responsiveness to disease progression in MCI than the ADAS-Cog 11, but were not better than the previously developed ADAS-Tree. Out of all measures

assessed, the ADAS-Plus EF&FA appeared to be the most responsive to treatment effects in MCI.

3.3.13 Common Item Pooling

Wouters et al. (2012) pooled data from 1863 subjects (585 NC, 66 MCI, 1012 AD, 133 non-AD dementia, 67 unknown psychiatric diagnosis) across multiple data sets which each included some or all of the CAMCOG, modified ADAS-Cog (ADAS-Cog 12 plus a concentration task), and MMSE.²⁰ Data pooling was performed using a method of “common item equating”, and Rasch measurement models were used to estimate the difficulty of each test item and the cognitive ability of each participant.²⁰ The purpose was to locate an underlying dimension of cognitive ability common to all three outcome measures so that their scores could be compared (the score from any one test can be translated to the level of underlying cognitive ability, and then translated back into a score on one of the other tests).²⁰ Items showing systematic differences in level of difficulty between data sets, or for which valid estimates of difficulty level could not be obtained, were excluded from common item pooling.²⁰

Rasch measurement theory was also applied to assess whether adding neuropsychological tests of episodic or semantic memory and executive function to the modified ADAS-Cog, CAMCOG, and MMSE increased precision for discriminating between levels of early cognitive decline and detecting mild dementia.²⁰ Neuropsychological tests were found to be more difficult than the modified ADAS-Cog, MMSE, and CAMCOG items with difficulty levels compatible with NC to MCI and mild dementia.²⁰ In contrast, the modified ADAS-Cog had only a few tasks with difficulty levels appropriate for pre-dementia cognitive abilities.²⁰

Baseline discrimination. The measurement precision for assessing levels of latent cognitive ability varied between the individual outcome measures as well as between different combinations of the outcome measures. For subjects with below average levels of cognitive ability, adding the MMSE and modified ADAS-Cog together (T-score range 50 to 60) improved precision for estimating underlying cognitive ability over either test alone.²⁰ For subjects with above average cognitive ability, adding neuropsychological

tests to the MMSE was the best measurement combination (T-score range 50 to 60).²⁰ The estimated difficulty level of the neuropsychological tests (T-score range 50-65) was more comparable with NC and MCI or mild dementia levels of cognitive ability than from the estimated difficulty level of the CAMCOG, MMSE, or modified ADAS-Cog tasks.²⁰ The neuropsychological tests were not however helpful for more severely impaired populations either alone, or when added to another measure.²⁰

Summary. The CAMCOG, MMSE, and modified ADAS-Cog estimate a common underlying dimension of cognitive ability. At mild levels of cognitive impairment, adding neuropsychological tests to the MMSE without the modified ADAS-Cog was recommended to maximize measurement precision, but for more severe levels of cognitive impairment adding the modified ADAS-Cog to the MMSE is advantageous over the modified ADAS-Cog alone. Formal assessments of responsiveness or comparisons to the ADAS-Cog 11 were not performed.

3.3.14 Alzheimer's Disease Neuroimaging Initiative Memory Composite

Crane et al. (2012) used modern psychometric approaches to develop and test the validity of a composite score for memory (ADNI Memory Composite) made up of the Rey Auditory Verbal Learning Test (RAVLT), ADAS-Cog 13 Delayed Word Recall task, ADAS-Cog 11 Word Recognition task, MMSE three word memory task with distractors, and the Logical Memory test which involves attempting to recall facts from a passage.²⁵ Initial analyses of the ADNI Memory Composite involved 225 subjects with NC, 394 subjects with MCI, and 184 subjects with AD.²⁵ Psychometric approaches determined that a bi-factor model was not a substantially better fit than a single factor model for the ADNI Memory Composite, so a single factor model was maintained.²⁵ Concurrent criterion validation found the ADNI Memory Composite performed at least as well as the RAVLT in all analyses.²⁵

Baseline discrimination. The ADNI Memory Composite score was slightly higher for subjects with NC (mean=1.0 points, SD=0.5) than subjects with MCI (mean=-0.1 points,

SD=0.6) and subjects with AD (mean=-0.8 points, SD=0.5).²⁵ No statistical tests of these differences were performed.

Disease progression. Ability of the ADNI Memory Composite to detect change over time in subjects with NC, MCI, and AD was evaluated using standardized regression coefficients for time, controlling for age, education, and sex, and presence of at least one APOEε4 allele.²⁵ Coefficients for time were statistically significant for the ADNI Memory Composite in NC (3.02), MCI (-9.43), and AD (-11.59) subgroups (all $P<0.05$).²⁵ In comparison, coefficients for the ADAS-Cog 11 were larger in the NC (3.20), MCI (-10.78), and AD (-12.25) subgroups (all $P<0.05$).²⁵

Treatment effect. Standardized coefficients and adjusted SD were used to estimate the sample size needed to detect a 25% reduction in rate of cognitive decline over 12 months with 80% power in a hypothetical two-arm clinical trial.²⁵ The ADNI Memory Composite required a smaller sample size than the ADAS-Cog 11 for a hypothetical trial of NC (28,512 versus 37,971), but required a larger sample size than the ADAS-Cog 11 for MCI (2,167 versus 1,651) and AD trials (568 versus 242).²⁵

Summary. The ADNI Memory Composite appeared able to discriminate between NC, MCI, and AD diagnostic categories. Although it demonstrated responsiveness to disease progression in NC, MCI, and AD samples, this performance was not better than that of the ADAS-Cog 11. The ADNI Memory Composite was more responsive to treatment effects in subjects with NC than the ADAS-Cog 11, but not more responsive to treatment effects for MCI and AD levels of disease severity.

3.3.15 ADAS-Cog IRT

Balsis et al. (2012) developed an IRT scoring methodology for the ADAS-Cog 11 in 1,240 subjects with varying levels of dementia severity.⁸³ Although the primary focus was on identifying limitations to traditional ADAS-Cog 11 scoring methodology, they also showed how using IRT to model a subject's score along with the difficulty of individual items can increase precision for estimating cognitive ability.⁸³

Verma et al. (2015) found, using multidimensional IRT on data from three cohort studies, that the ADAS-Cog 11 is most appropriately modelled using three latent factors corresponding to the cognitive domains of memory, language, and praxis in a large sample of older adults with NC to AD.⁸⁴ The memory domain includes the Word Recall, Orientation, and Word Recognition tasks.⁸⁴ The Language domain includes the Naming Objects and Fingers, Language, Comprehension of Spoken Language, Word Finding Difficulty, and Remembering Test Instructions tasks.⁸⁴ The praxis domain includes the Commands, Constructional Praxis, and Ideational Praxis tasks.⁸⁴

Verma et al. (2015) evaluated their multidimensional IRT scoring methodology for the ADAS-Cog 11 using data from the treatment arms of 11 clinical trials that enrolled older adults with AD. Their ADAS-Cog IRT uses ICCs from patient responses on the ADAS-Cog 11 to provide an assessment of cognitive impairment based on maximum likelihood estimation.⁸⁴ Differential Item Functioning (DIF) analyses were used to adjust item slopes and intercepts so that patient characteristics other than cognitive ability did not cause large variations in scores.⁸⁴ To maintain non-negative integer final scores, the summary scores for memory, language, and praxis domains were linearly scaled by multiplying by a factor of 15 and adding 50.⁸⁴ This ADAS-Cog IRT scoring methodology demonstrated good accuracy as assessed by root mean squared error of observed compared to predicted ADAS-Cog 11 scores (6.05 points).⁸⁴ Precision was assessed using item information functions. Memory tasks showed good precision across the entire range of memory impairment, however precision for measuring language and praxis impairment was only good at lower levels of cognitive ability.⁸⁴

Baseline discrimination. The application of IRT scoring methodology to the ADAS-Cog 11 provided the same score to all subjects with the same cognitive ability.⁸³ In contrast, it was found that when using traditional ADAS-Cog 11 scoring methodology two subjects with the same cognitive ability may score differently, and two subjects with different scores on the ADAS-Cog 11 may have the same underlying cognitive ability.⁸³

Treatment effect. Verma et al. (2015) used clinical trial simulations to compare the ADAS-Cog 11 and ADAS-Cog IRT in terms of the power needed to detect a pre-specified treatment effect for various sample sizes (n=200 to 1,000) over 24 months, and

for various lengths of follow-up with the sample size set at 400.⁸⁴ Both ADAS-Cog IRT scoring methodology and original ADAS-Cog 11 scoring with an Analysis of Covariance test for a treatment effect showed low power (< 80%) for detecting a mild treatment effect regardless of the sample size or trial duration.⁸⁴ For a moderate treatment effect, ADAS-Cog IRT methodology reached 80% power with a smaller sample size and shorter trial duration compared to original ADAS-Cog 11 methods.⁸⁴ Sensitivity analysis in a real clinical trial was also performed where the ADAS-Cog IRT scoring methodology detected a larger treatment effect than original ADAS-Cog 11 methods.⁸⁴

Summary. The ADAS-Cog IRT demonstrated more precise estimates of cognitive ability than original ADAS-Cog 11 scoring methodology, which is expected to improve responsiveness to baseline discrimination, and the ADAS-Cog IRT demonstrated greater responsiveness to moderately large treatment effects in AD. The finding that the ADAS-Cog 11 was best modelled using multiple latent cognitive domains suggests that the unidimensional assumption used in CTT may not be appropriate for assessing the ADAS-Cog 11. Responsiveness to disease progression was not evaluated.

3.3.16 ADAS-3

Raghavan et al. (2013) aimed to improve sensitivity to change and reduce variability of the ADAS-Cog 11 for MCI and early AD trials by removing uninformative items from the ADAS-Cog 11 and adding in more responsive measures of cognition or function.¹⁹ A total of six novel measures were derived based on analyses of cognitive and functional measures in 229 subjects with NC, 377 subjects with MCI, and 192 subjects with AD.¹⁹ The criterion for an individual test item to be considered for inclusion in a novel composite was a standardized two-year change score of at least 0.4 for MCI participants.¹⁹ Three of the novel measures were composed solely of cognitive test items (Section 3.3.16 and 3.3.17), and three included cognitive items as well as measures of daily function (Section 3.3.18). Bootstrap validation was performed for the entire selection process. Performances of the six novel measures for detecting change over time were compared with each other as well as with other outcome measures, including the ADAS-Cog 11 and ADAS-Cog 13, using data from two-years of follow-up of 198

subjects with NC, 138 subjects with stable MCI, 139 subjects who converted from MCI to dementia, and 131 subjects with AD.¹⁹

The first cognitive measure, the ADAS-3, includes ADAS-Cog 11 tasks which did not exhibit ceiling effects and surpassed the 0.4 threshold for standardized two-year change scores. ADAS-3 tasks include: Word Recall, Delayed Word Recall, and Orientation.¹⁹

Disease progression. The standardized two-year change of the ADAS-3 was larger than that of the ADAS-Cog 11 and all individual tasks of the ADAS-Cog 11, but smaller than that of the other five novel composites.¹⁹

Summary. The ADAS-3 was more responsive to disease progression than the ADAS-Cog 11 among a sample of subjects with NC to AD levels of disease severity, but it was the worst performing novel composite developed by Raghavan et al (2013).

Responsiveness to baseline discrimination and treatment effects was not evaluated.

3.3.17 Cognitive Composites 1 and 2

The second novel composite developed by Raghavan et al. (2013), the Cognitive Composite (CC) 1 (CC1), includes the same items as the ADAS-3 as well as the RAVLT immediate recall test, and the MMSE.¹⁹

The third novel composite developed by Raghavan et al. (2013), the CC2, consists of the ADAS-3 and the cognitive portion of the Clinical Dementia Rating Scale – Sum of Boxes (CDR-SB).¹⁹

Criterion validation was performed for both the CC1 and CC2 using Spearman's correlations between two-year change scores for each of the composites and reference standards such as the ADAS-Cog 11 and CDR-SB, and factor analysis was used to assess the latent structure of each novel composite measure.¹⁹

Baseline discrimination. For the CC1, subjects with MCI scored worse (mean=0.15 points, SD=1.64) than subjects with AD (mean=3.15 points, SD=1.68), or subjects with MCI and A β pathology (mean=0.49 points, SD=1.55).¹⁹ The same was true for the CC2,

whereby subjects with MCI scored worse (mean=0.07 points, SD=0.94) than subjects with AD (mean=2.38 points, SD=1.28), or subjects with MCI and A β pathology (mean=0.22 points, SD=0.95).¹⁹ Statistical significance of these differences was not assessed.

Disease progression. The CC1 and CC2 demonstrated greater standardized two-year mean change than the ADAS-Cog 11 and all individual items of the ADAS-Cog 11, and the ADAS-3.¹⁹ The CC2 was the most responsive purely cognitive measure developed by Ragavan et al. (2013), producing a standardized two-year change score only slightly smaller than the best performing composite incorporating items of cognition and functionality.¹⁹ Two-year change scores from the CC1 were more strongly correlated with ADAS-Cog 11 (Spearman's rho=0.61) than CC2 with the ADAS-Cog 11 (Spearman's rho=0.54).¹⁹

Treatment effect. The estimated sample size required to detect a hypothetical 25% treatment effect with 80% power in a two-arm clinical trial of subjects with MCI was smaller for the CC2 (n=300) than the CC1 (n=477), the ADAS-Cog 11 (n=772), and the CDR-SB (n=375).¹⁹

Summary. The CC1 and CC2 both appeared able to discriminate between groups of subjects with MCI and AD, and were more responsive than the ADAS-Cog 11 to disease progression in subjects with NC to AD. Results also suggest that both the CC1 and CC2 are more responsive to treatment effects in subjects with MCI than the ADAS-Cog 11. The CC2 demonstrated similar responsiveness to the novel composites which include functional measures, while maintaining lower variability.

3.3.18 Cognitive Functional Composites 1 and 2

The fourth novel measure developed by Raghavan et al. (2013), the Cognitive Functional Composite (CFC) 1 (CFC1), was the first of their three novel composites which included measures to assess both cognition and daily functioning.¹⁹ The CFC1 is composed of the CC1 and the FAQ.¹⁹

The fifth composite derived by Raghavan et al. (2013), the CFC2, includes the CC2 and the FAQ.¹⁹

The sixth composite was the only novel measure derived by Raghavan et al. (2013) which did not incorporate any ADAS-Cog 11 items (CFC3: CDR-SB and FAQ).¹⁹ For that reason, it will not be reviewed in the same manner as the other novel composites.

Correlations between the cognitive portions of each of the CFCs with the FAQ were used to demonstrate that change scores on the CFCs were due to changes on both cognitive and functional sub-tasks rather than just being driven by one of the two domains.¹⁹

Baseline discrimination. For the CFC1, subjects with MCI scored worse (mean=-0.11 points, SD=1.02) than subjects with AD (mean=2.4 points, SD=1.42), or subjects with MCI and A β pathology (mean=0.06 points, SD=0.98).¹⁹ The same was true for the CFC2, whereby subjects with MCI scored worse (mean=-0.13 points, SD=1.0) than subjects with AD (mean=2.48 points, SD=1.51), or subjects with MCI and A β pathology (mean=0 points, SD=1.01).¹⁹ Statistical significance of these differences was not assessed.

Disease progression. Among MCI participants, the two-year standardized mean change of the CFC1 and CFC2 were larger than that of the ADAS-Cog 11, the CDR-SB, and all individual items of the ADAS-Cog 11.¹⁹ The CFC2 demonstrated the largest standardized mean change of all novel measures developed by Raghavan et al. (2013).¹⁹ Spearman's correlation between two-year change scores on the CFC1 and on the ADAS-Cog 11 ($\rho=0.54$) was slightly higher than for the CFC2 and the ADAS-Cog 11 ($\rho=0.48$).¹⁹

Treatment effect. The estimated sample size required to detect a hypothetical 25% treatment effect with 80% power in a two-arm clinical trial of subjects with MCI was smaller for the CFC2 (n=302) than the CFC1 (n=348), the ADAS-Cog 11 (n=772), and the CDR-SB (n=375).¹⁹

Summary. The CFC1 and CFC2 appeared able to discriminate between MCI and AD groups, and both were more responsive than the ADAS-Cog 11 to disease progression in subjects with NC to AD. It was also suggested that both the CFC1 and CFC2 would be

more responsive than the ADAS-Cog 11 to treatment effects in MCI. Overall, the CFC2 was the most responsive measure developed by Raghavan et al. (2013).

3.3.19 Item Response Theory and Pharmacometric ADAS-Cog 13

Ueckert et al. (2014) used IRT and pharmacometric modelling to explore different methods for analyzing ADAS-Cog 13 scores.¹⁵

First, an IRT model to estimate baseline cognitive ability was created using data from 2,744 subjects with NC, MCI, or mild AD.¹⁵ The *IRT baseline model* models cognitive ability as a subject specific random effect following a standard normal distribution (Z-score), with no limits on the upper or lower extremes of cognitive ability.¹⁵ The probability of a subject responding a certain way on an ADAS-Cog 13 task or task subitem, given their underlying cognitive ability, was described using four different test item specific models. First, tasks or subitems that are scored as correct or incorrect (e.g. Orientation subitem: correctly state the month) were modeled with a *three-parameter binary model* that accounts for item discrimination, item difficulty, and the probability that a subject with no cognitive disability would get the item incorrect.¹⁵ Second, tasks or task subitems involving words were modeled with a *binomial model* (uses the number of words correctly identified out of the total number possible), with slightly different failure probabilities depending on the task (Word Recall: failure probability = three-parameter binary model described above; Word Recognition: failure probability = same as for Word Recall plus a fourth parameter to account for the maximal probability that a subject with severe cognitive dysfunction would correctly categorize words as seen or not).¹⁵ All words were assumed to hold the same amount of information about underlying cognition.¹⁵ Third, the Number Cancellation task was modelled using a *generalized Poisson model*, which included the same three test item parameters as the three-parameter binary model plus a fourth parameter for dispersion, and a factor to ensure predicted scores are in the range of 0 to 40.¹⁵ Fourth, tasks on the ADAS-Cog 13 that are rated by the examiner (e.g. Comprehension of Spoken Language) were modeled using a *proportional odds, ordered categorical model* with five possible categories (none to severe impairment) and parameters for item difficulty and discrimination.¹⁵

Next, three different means for assessing cognitive change over time were devised. The latter two are extensions of the baseline IRT model:

- i) A *Least-Square Mean Analysis Model* used change in ADAS-Cog 13 score as the outcome variable, treatment as the exposure variable, visit as a repeated factor, baseline ADAS-Cog 13 score as a covariate, a treatment-by-visit interaction term, and a grouping factor of subjects nested within treatment. This represents more “traditional” ADAS-Cog scoring methodology.¹⁵
- ii) The baseline IRT model was extended to create a *Longitudinal IRT Model* by adding a hidden variable to account for disease progression over time.¹⁵ Disease progression was assumed to be linear (based on a previously published model), subject-specific, and modelled through random-effects. A hazard function for the probability that a subject will drop out of a longitudinal study was also included.¹⁵ To assess the performance of the longitudinal IRT model, Z-score estimates of underlying cognitive ability for an 18-month long clinical trial were translated back to the original ADAS-Cog 13 scoring scale and compared with observed ADAS-Cog 13 scores from 322 real subjects in the 18-month clinical trial.¹⁵ More specifically, two-hundred Monte-Carlo simulations from the IRT model and the original clinical trial data were used to compare the proportion of subjects from the original data whose task-level scores would fall in the 95% prediction interval from the score produced by IRT models.¹⁵ Total ADAS-Cog 13 score comparisons were done in a similar manner, except 200 non-Bayesian simulations were performed and the 95% CI for the median, 2.5th, and 97.5th percentile of total simulated scores were compared to the clinical trial percentiles.¹⁵ ADAS-Cog 13 scores for the clinical trial were plotted with the median, 2.5th, and 97.5th percentile of the real scores observed in the clinical trial. When plotted on top, the median value of the real scores fell within the 95% CI predicted by the IRT model for all but the final 18-month follow-up assessment.¹⁵ Comparisons were also made for drop-out patterns over the course of the clinical trial.¹⁵

- iii) The *Pharmacometric Total ADAS-Cog Score Model* of analysis was based on a previously published disease progression model and modified according to the results of goodness of fit plots, residual plots, and visual prediction checks.¹⁵ This model was further refined and tested using a simulated data set from the longitudinal IRT model.¹⁵ Similar to the longitudinal IRT model analyses whereby estimated ADAS-Cog 11 total scores were compared to observed scores in a real clinical trial, the performance of the pharmacometric total ADAS-Cog 13 score model was assessed with visual predictive checks of whether the 95% CI for the ADAS-Cog 13 scores estimated from the pharmacometric model included the 2.5th, 97.5th, and median ADAS-Cog 13 scores from the ADAS-Cog Longitudinal IRT model based simulated data set.¹⁵ The final pharmacometric total ADAS-Cog 13 score model assumes a linear progression of cognitive dysfunction (increasing scores), and models individual subject baseline scores with a Box-Cox distribution and normally distributed individual slope parameters correlated with baseline random effect.¹⁵

Treatment effect. The longitudinal IRT model was used to simulate 20-month two-arm clinical trials with a 20% treatment effect for 100, 200, 400, or 800 subjects with mild to moderate AD.¹⁵ Five hundred simulations were run for each sample size.¹⁵ Type I error and power to detect the treatment effect of the three different methods of longitudinal data analysis described above were compared.¹⁵ The IRT based pharmacometric model required 71% fewer subjects than the Least-square mean analysis, and 23% fewer subjects than the pharmacometric model, to detect a treatment effect with 80% power and no inflation of Type I error.¹⁵

Summary. Using both IRT and pharmacometric modelling demonstrated greater precision of cognitive ability estimates at baseline, and appeared more responsive to treatment effects in AD, compared to traditional ADAS-Cog scoring and methods of analysis. Responsiveness to baseline discrimination and disease progression was not assessed.

3.3.20 integrated Alzheimer's Disease Rating Scale

Wessels et al. (2015) used a theoretical framework to guide the combination of existing scales of cognition and function to create a sensitive measure to the natural history of AD and to detect treatment effects in clinical trials.⁸⁵ For subjects with mild AD and MCI, preliminary tests found the combination of ADAS-Cog 13 and the FAQ was most sensitive, and the combination of the two scales performed better than either one individually.⁸⁵ Data from treatment trials in AD did not have those two measures specifically, so they were approximated with the ADAS-Cog 14 and the ADCS-instrumental Activities of Daily Living (iADL) which formally make up the integrated Alzheimer's Disease Rating Scale (iADRS).⁸⁵

iADRS score = $[-1(\text{ADAS-Cog } 14) + 90] + \text{iADL}$, where the ADAS-Cog 14 (ADAS-Cog 11, Delayed Word Recall, Maze, and Digit Cancellation tasks) and iADL are summed normally and the total range of the iADRS is 0 to 146 with lower scores indicating worse performance.⁸⁵

Psychometric analyses showed that the iADRS is composed of two principal components (cognition and instrumental function) for assessment at one point in time, and the majority of the variability for subjects with MCI was due to cognitive items of the ADAS-Cog.⁸⁵ For change over time, the iADRS items load on a single component, and variance of change scores was driven by both cognitive and function items.⁸⁵

Disease progression. SRMs with 95% CIs were compared using separate forest plots for different levels of disease severity. The iADRS had the largest SRM for MCI and mild and moderate AD compared to the ADAS-Cog 11, ADAS-Cog 13, MMSE, FAQ, CDR-SB, and several other measures of cognition.⁸⁵

Treatment effect. For several clinical trials including subjects with MCI or mild AD the iADRS was able to detect a statistically significant treatment effect, however the magnitude of this effect was not consistently better than that detected by the ADAS-Cog 14.⁸⁵

Summary. The iADRS was more responsive to disease progression in MCI and AD than the ADAS-Cog 11, however it was not more responsive than the ADAS-Cog 14 for treatment effects in MCI and mild AD. Comparison of responsiveness to treatment effects with the ADAS-Cog 11 was not performed, and baseline discrimination ability was not evaluated.

3.3.21 Straightforward Sensitive Scale

Huang et al. (2015) designed a scale including cognitive and functional measures for the purpose of tracking disease progression over time and detecting potential treatment effects in clinical trials for MCI and early AD, while maintaining good reliability and validity as subjects progress to more severe stages of AD.⁸⁶ Selection of measures to include in the composite scale was performed in a stepwise manner. First, SRMs of many candidate measures were calculated and the candidate measures with the highest SRMs were combined to create a composite measure.⁸⁶ In general, the minimum SRM for a candidate measure to be considered was 0.45 for the group of 397 subjects with MCI, 0.50 for an APOE enriched subgroup, and 0.55 for hippocampal volume and A β enriched subgroups.⁸⁶ The SRMs of all possible combinations of candidate measures were calculated to determine the composite scale most sensitive to disease progression and treatment effects.⁸⁶ This “straightforward sensitive scale” (SSS) consisted of the CDR-SB, FAQ, and three ADAS-Cog 13 items (Word Recall, Delayed Word Recall, Orientation).⁸⁶

Disease progression. The SRM of the SSS in subjects with MCI was greater than that of the CDR-SB alone or the ADAS-Cog 13 over 1 year (SRM: SSS=0.62, CDR-SB=0.55, ADAS-Cog 13=0.28), two years (SRM: SSS=0.82, CDR-SB=0.74, ADAS-Cog 13=0.56), three years (SRM: SSS=0.93, CDR-SB=0.76, ADAS-Cog 13=0.65), and when assuming a hypothetical treatment effect delayed disease progression by one year (SRM: SSS=0.37, CDR-SB=0.35, ADAS-Cog 13=0.29).⁸⁶ The SSS maintained the highest SRMs for subgroups of subjects with MCI and biomarkers indicating increased risk of disease progression.⁸⁶

Treatment effect. The SSS was estimated to require a smaller sample size (n=189) to detect a hypothetical treatment effect that slows disease progression by 50% in a two-year MCI trial compared to the CDR-SB (n=231) and ADAS-Cog 13 (n=402).⁸⁶

Summary. The SSS appeared more responsive to disease progression and treatment effects in MCI than the ADAS-Cog 13, but direct comparison with the ADAS-Cog 11 was not performed nor were tests of baseline discrimination.

3.3.22 ADAS-Cog 3b

Podhorna et al. (2016) removed eight tasks from the ADAS-Cog 11 that demonstrate ceiling effects in MCI. The three remaining tasks comprise the ADAS-Cog 3b: Word Recall, Orientation, and Word Recognition.¹² The ADAS-Cog 3b assesses only memory and has a scoring range of 0 to 30.¹²

Baseline Discrimination. ADAS-Cog 3b scores for 382 subjects with MCI (mean=8.23 points, SD=3.76) were on average lower than those of 97 subjects with mild AD (mean=15.95 points, SD=4.15).¹² Podhorna et al. (2016) further divided the 382 subjects with MCI into two groups depending on whether CSF and APOEε4 biomarkers of AD pathology were present. The enriched MCI subgroup (n=206) had worse scores at baseline (mean=9.43 points, SD=3.92) on the ADAS-Cog 3b than the non-enriched MCI subgroup (n=176, mean=6.82 points, SD=3.02).¹² Podhorna et al. (2016) also found scores on the ADAS-Cog 11 were lower in MCI (mean=9.50 points, SD=4.29) than mild AD (mean=19.66 points, SD=6.30) groups, and in the non-enriched (mean=7.94 points, SD=3.50 points) compared to the enriched MCI subgroup (mean=10.83, SD=4.46).¹² Tests of statistical significance for these differences were not performed.

Disease progression. Podhorna et al. (2016) found very little change on the ADAS-Cog 3b in 382 subjects with MCI over 24 months (mean=0.71 points, SD=3.56) and in 169 subjects with MCI over 36 months (mean=1.23 points, SD=4.00).¹² There was very little change in the enriched MCI subgroup (mean 24 month change=1.48 points, SD=3.78; mean 36 month change=2.55 points, SD=4.40), and almost no change in the non-enriched subgroup (mean 24 month change=-0.19 points, SD=3.06; mean 36 month change=-0.25

points, SD=5.12).¹² Among 97 subjects with mild AD there was also very little change in mean ADAS-Cog 3b score over 12 months (mean=1.82 points, SD=3.91) and 24 months (mean=3.81 points, SD=5.12).¹² The SRM for change over 24 months in 382 subjects with MCI was 0.42 (95% CI 0.20, 0.61) for the ADAS-Cog 3b and 0.37 (95% CI 0.15, 0.57) for the ADAS-Cog 11, adjusting for baseline MMSE, age, sex, APOEε4 allele.¹² The SRM for change over 12 months in 97 subjects with mild AD was 0.81 (95% CI 0.43, 1.09) for the ADAS-Cog 3b and 0.87 (95% CI 0.46, 1.13) for the ADAS-Cog 11, adjusting for baseline MMSE, age, sex, APOEε4 allele.¹² SRMs were not statistically different from each other (all $P>0.10$).¹²

Summary. The ADAS-Cog 3b appeared able to discriminate between groups of subjects with MCI and mild AD, however responsiveness to disease progression in MCI or mild AD was not superior to the ADAS-Cog 11. Responsiveness to treatment effects was not assessed.

3.3.23 ADAS-Cog 5

Podhorna et al. (2016) created the ADAS-Cog 5 by adding to the ADAS-Cog 3b Delayed Word Recall, and Digit Cancellation tasks.¹² The additional tasks assess attention and executive function, and the ADAS-Cog 5 is scored from 0 to 45.¹²

Baseline Discrimination. ADAS-Cog 5 scores for 382 subjects with MCI (mean=13.96 points, SD=6.17) were lower than those of 97 subjects with mild AD (mean=26.20 points, SD=5.31).¹² The 382 subjects with MCI were further divided into two groups depending on whether CSF and APOEε4 biomarkers of AD pathology were present.¹² The enriched MCI subgroup (n=206) had worse scores on the ADAS-Cog 5 at baseline (mean=16.12 points, SD=6.28) than the non-enriched subgroup (n=176, mean=11.43 points, SD=4.99).¹² Statistical significance of the above differences was not tested.

Disease progression. Podhorna et al. (2016) found almost no change on the ADAS-Cog 5 in 382 subjects with MCI over 24 months (mean change=1.13 points, SD=4.87) or in 168 subjects with MCI over 36 months (mean change=1.95 points, SD=5.58).¹² There was very little difference in the enriched MCI subgroup (mean 24 month change=2.21

points, SD=5.58; mean 36 month change=3.82 points, SD=6.03), and in the non-enriched subgroup scores there was no meaningful change (mean 24 month change=-0.11 points, SD=4.12; mean 36 month change=-0.16 points, SD=4.15).¹² Among 97 subjects with mild AD there also was very little change on the ADAS-Cog 5 score over 12 months (mean change=2.64 points, SD=4.39) and 24 months (mean change=5.48 points, SD=6.13).¹² The SRM for the ADAS-Cog 5 for 382 subjects with MCI over 24 months was 0.42 (95% CI 0.19, 0.63), adjusting for baseline MMSE, age, sex, APOEε4 allele.¹² The SRM for change on the ADAS-Cog 5 over 12 months in 97 subjects with mild AD was 0.93 (95% CI 0.52, 1.22), adjusting for baseline MMSE, age, sex, APOEε4 allele.¹² SRMs for the ADAS-Cog 5 were not significantly different than SRMs for the ADAS-Cog 11, ADAS-Cog 13, or ADAS-Cog 3b (all $P>0.10$).¹²

Summary. The ADAS-Cog 5 appeared able to discriminate between groups of subjects with MCI and mild AD, however responsiveness to disease progression in MCI was not superior to the ADAS-Cog 11. Responsiveness to treatment effects was not evaluated.

3.3.24 ADAS-13 Re-Weighted

Grochowalski et al. (2016) created three different versions of the ADAS-Cog 13 (Section 3.3.1) using data from 153 subjects with AD and 352 subjects with MCI in an effort to improve reliability of ADAS-Cog 13 change scores.⁸⁷ Improved reliability would reduce variability and ultimately improve the ability of the ADAS-Cog 13 to track changes in cognition over time. The three versions included a re-weighted ADAS-Cog 13, a lengthened ADAS-Cog 13, and a re-weighted and lengthened ADAS-Cog 13.⁸⁷ To obtain these three different variants the ADAS-Cog 13 was divided into three subsections based on task scoring procedures.⁸⁷ Each section was given a separate weight, calculated as the number of tasks in that section divided by the total number of tasks in the test.⁸⁷ This resulted in a section of verbal memory with weight 0.10, a section of clinician-rated tasks with weight 0.45, and a section for general cognitive tasks with weight 0.45.⁸⁷ Lengthening the test, either with or without re-weighting, did not substantially improve score reliability so the authors concluded that the ADAS-Cog 13 with re-weighted sections was the best variant for improving reliability of change scores (ADAS-13RW).⁸⁷

Criterion validation was performed by analysing correlations between the ADAS-13RW and the ADAS-11, ADAS-13, RAVLT, and MMSE.⁸⁷

Disease progression. Reliability of change scores of the ADAS-13RW was better than the ADAS-Cog 13, but only of an “acceptable” magnitude for change scores defined by cut-score dependability (compare subject’s scores to pre-set criterion value of 4 points change).⁸⁷ Relative change score reliability (rank subject’s change relative to another subject’s change) and absolute change score reliability (estimate of subject’s true individual change score) were not of an acceptable magnitude.⁸⁷

Summary. Re-weighting and/or lengthening the ADAS-Cog 13 did not improve the reliability of change scores for MCI to a level recommended for assessing meaningful clinical change. Assessments of responsiveness were not performed.

3.3.25 Alzheimer’s Disease Composite Score

Wang et al. (2016) developed the Alzheimer’s Disease Composite Score (ADCOMS) from outcome measures previously shown to be sensitive to AD-specific clinical decline and treatment effects in subjects with MCI.⁸⁸ A partial least squares procedure was used to fit a linear model characterizing disease progression and variable importance projections (VIP) for numerous candidate items.⁸⁸ The ADCOMS was derived by combining the twelve items that demonstrated a VIP value of 0.8 or greater using their partial least squares coefficients as a weighting factor.⁸⁸ Specific items included four ADAS-Cog 12 tasks (Delayed Word Recall, Orientation, Word Recognition, Word Finding Difficulty), two MMSE items (Orientation time, Drawing), and six CDR-SB items (Personal Care, Community Affairs, Home and Hobbies, Judgement and Problem Solving, Memory, Orientation).⁸⁸

Disease progression. The 12-month SRM of the ADCOMS (0.419) was larger than that of the ADAS-Cog 12 (0.196), MMSE (0.221), and CDR-SB (0.353) for a pooled aMCI sample, as well as subgroups of aMCI subjects with genetic or CSF AD biomarkers present.⁸⁸

Treatment effect. The ADCOMS was able to detect a statistically significant treatment effect for donepezil compared to placebo for aMCI participants ($P=0.02$), which was also found by the MMSE ($P=0.02$), but not by the ADAS-Cog 12 ($P=0.12$) or CDR-SB ($P=0.11$).⁸⁸ The ADCOMS did not find a statistically significant effect for vitamin E in subjects with aMCI ($P=0.89$), nor did the ADAS-Cog 12 ($P=0.76$), MMSE ($P=0.59$), or CDR-SB ($P=0.42$).⁸⁸ The ADCOMS was also able to detect a statistically significant treatment effect for donepezil in subjects with mild AD ($P<0.0001$) as did the ADAS-Cog 12 ($P=0.0008$), MMSE ($P=0.001$), and CDR-SB ($P=0.02$).⁸⁸

Summary. The ADCOMS demonstrated better responsiveness to disease progression and treatment effects in MCI than the ADAS-Cog 12, but the ADAS-Cog 11 was not analyzed. Tests of baseline discrimination were also not performed.

3.3.26 Summary of Modifications Made to the ADAS-Cog 11

A total of thirty-one modifications of the ADAS-Cog 11 were found. Five of these modifications altered the scoring methodology for the original ADAS-Cog 11, four maintained the original scoring methodology and added additional tasks, and twenty-two altered both scoring methodology and included additional item content.

Results from studies which compared the modified outcome measure to the ADAS-Cog 11 suggested that 13 modification approaches demonstrated responsiveness to group-level between person differences in stage of disease progression at one point in time (baseline discrimination), seven improved responsiveness to group-level within-person observed change measured over time (disease progression; natural history) for MCI samples, three were more responsive to this type of change than the ADAS-Cog 11 for dementia samples, and five were more responsive to this type of change than the ADAS-Cog 11 in samples with various levels of cognitive ability (mixed dementia and pre-dementia syndromes). One modification was found to improve responsiveness to group-level between-person differences of within-person observed change over time (treatment effect) in subjects with NC, ten were more responsive to treatment effects in subjects with MCI, and six were more responsive to treatment effects in subjects with dementia than the ADAS-Cog 11. It is possible that several of the other modified versions of the ADAS-

Cog 11 also improved performance, but these have not yet been evaluated or compared to the ADAS-Cog 11. In general, the CV (Section 2.2.5 for more information on CV) for the ADAS-Cog 11 and modified versions was reduced as disease severity worsened, and several of the modifications reduced the CV within each diagnostic category of NC, MCI, and dementia (Appendix A).

3.4 Summary of Chapter 3

The ADAS-Cog 11 was developed to assess cognitive dysfunction in moderate to severe AD. Since the time of development there has been a shift in the field of dementia research towards studying pre-dementia syndromes. The ADAS-Cog 11 continues to be used for these pre-dementia studies, however its performance is limited due to ceiling effects, suboptimal scoring methodology, and poor content validity. Modifying the scoring methodology of the ADAS-Cog 11 improves its responsiveness to several types of change, as does adding additional item content. In particular, tasks assessing EF, delayed recall, and daily functioning, which are now known to be important components of pre-dementia disease severity and progression, improve the content validity of the ADAS-Cog 11 as well as its responsiveness. None of the modifications of the ADAS-Cog 11 included assessments of motor performance, which has also been shown to be an important component of pre-dementia syndromes and disease progression.

3.5 References

1. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's Disease. *Am J Psychiatry*. 1984;141(11):1356-1364. doi:10.1176/ajp.141.11.1356.
2. Zec RF, Landreth ES, Vicari SK, et al. Alzheimer Disease Assessment Scale: useful for both early detection and staging of dementia of the Alzheimer types. *Alzheimer Dis Assoc Disord*. 1992;6(2):89-102.
3. Skinner J, Carvalho JO, Potter GG, et al. The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): an expansion of the ADAS-Cog to improve responsiveness in MCI. 2012;6:489-501. doi:10.1007/s11682-012-9166-3.
4. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI) clinical characterization. *Neurology*. 2010;74:201-209.
5. Pyo G, Elble RJ, Ala T, Markwell SJ. The characteristics of patients with uncertain/mild cognitive impairment on the Alzheimer Disease Assessment Scale-Cognitive Subscale. *Alzheimer Dis Assoc Disord*. 2006;20(1):16-22. doi:10.1097/01.wad.0000201846.22213.76.
6. Mahoney R, Johnston K, Katona C, Maxmin K, Livingston G. The TE4D-Cog: a new test for detecting early dementia in English-speaking populations. *Int J Geriatr Psychiatry*. 2005;20:1172-1179. doi:10.1002/gps.1412.
7. Sano M, Raman R, Emond J, et al. Adding delayed recall to the Alzheimer Disease Assessment Scale is useful in studies of Mild Cognitive Impairment but not Alzheimer Disease. *Alzheimer Dis Assoc Disord*. 2011;25(2):122-127.
8. Zhang N, Song X, Zhang Y, Chen W, Arcy R. An MRI brain atrophy and lesion index to assess the progression of structural changes in Alzheimer's Disease, Mild Cognitive Impairment, and Normal Aging: a follow-up study. 2011;26:359-367. doi:10.3233/JAD-2011-0048.
9. Lam LCW, Tam CWC, Lui VWC, et al. Use of clinical dementia rating in detecting early cognitive deficits in a community-based sample of Chinese older persons in Hong Kong. *Alzheimer Dis Assoc Disord*. 2008;22(2):153-157.
10. Evans MC, Barnes J, Nielsen C, et al. Volume changes in Alzheimer's Disease and Mild Cognitive Impairment: cognitive associations. 2010;20:674-682. doi:10.1007/s00330-009-1581-5.
11. Steenland K, Zhao L, Goldstein F, Cellar J, Lah J. Biomarkers for predicting cognitive decline in those with normal cognition. *J Alzheimers Dis*. 2014;40(3):587-594. doi:10.3233/JAD-2014-131343.
12. Podhorna J, Krahnke T, Shear M, E Harrison J. Alzheimer's Disease Assessment Scale-Cognitive subscale variants in Mild Cognitive Impairment and mild Alzheimer's Disease: change over time and the effect of enrichment strategies. *Alzheimers Res Ther*. 2016;8(8):1-13. doi:10.1186/s13195-016-0170-5.
13. Schrag A, Schott JM, Alzheimer's Disease Neuroimaging I. What is the clinically relevant change on the ADAS-Cog? *J Neurol Neurosurg Psychiatry*. 2012;83:171-173. doi:10.1136/jnnp-2011-300881.
14. Zanutta D, Puricelli S, Bonoldi G. Cognitive effects of a dietary supplement made from extract of *Bacopa monnieri*, astaxanthin, phosphatidylserine, and vitamin E in subjects with Mild Cognitive Impairment: a noncomparative, exploratory clinical study. *Neuropsychiatr Dis Treat*. 2014;10:225-230. doi:10.2147/NDT.S51092.

15. Ueckert S, Plan EL, Ito K, Karlsson MO, Corrigan B, Hooker AC. Improved utilization of ADAS-Cog assessment data through Item Response Theory based Pharmacometric Modeling. *Pharm Res*. 2014;31(8):2152-2165. doi:10.1007/s11095-014-1315-5.
16. Graham DP, Cully JA, Snow AL, Massman P, Doody R. The Alzheimer's Disease Assessment normative data for older adult controls. 2004;18(4):236-240.
17. Grundman M, Petersen RC, Ferris SH, et al. Mild Cognitive Impairment can be distinguished from Alzheimer Disease and normal aging for clinical trials. *Arch Neurol*. 2004;61:59-66.
18. Llano D, Laforet G, Devanarayan V. Derivation of a new ADAS-Cog composite using tree-based multivariate analysis: prediction of conversion from Mild Cognitive Impairment to Alzheimer Disease. *Alzheimer Dis Assoc Disord*. 2011;25(1):73-84.
19. Raghavan N, Samtani MN, Farnum M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimer's Dement*. 2013;9(10):S21-S31. doi:10.1016/j.jalz.2012.05.2187.
20. Wouters H, Appels B, van der Flier WM, et al. Improving the accuracy and precision of cognitive testing in mild dementia. *J Int Neuropsychol Soc*. 2012;18(2):314-322. doi:10.1017/S1355617711001755.
21. Alexander GE, Satalich TA, Shankle WR, Batchelder WH. A cognitive psychometric model for the psychodiagnostic assessment of memory-related deficits. *Psychol Assess*. 2016;28(3):279-293. doi:10.1037/pas0000163.
22. Connor DJ, Sabbagh MN. Administration and scoring variance on the ADAS-Cog. *J Alzheimers Dis*. 2008;15(3):461-464.
23. Herholz K, Westwood S, Haense C, Dunn G. Evaluation of a calibrated (18)F-FDG PET score as a biomarker for progression in Alzheimer Disease and Mild Cognitive Impairment. *J Nucl Med*. 2011;52:1218-1226. doi:10.2967/jnumed.111.090902.
24. Farlow MR, He Y, Tekin S, Xu J, Lane R, Charles HC. Impact of APOE in Mild Cognitive Impairment. *Neurology*. 2004;63:1898-1901.
25. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav*. 2012;6:502-516. doi:10.1007/s11682-012-9186-z.
26. Daiello LA, Gongvatana A, Dunsiger S, Cohen RA, Ott BR. Association of fish oil supplement use with preservation of brain volume and cognitive function. *Alzheimers Dement*. 2015;11(2):226-235. doi:doi:10.1016/j.jalz.2014.02.005.
27. Doraiswamy PM, Sperling RA, Coleman RE, et al. Amyloid-Beta assessed by florbetapir F 18 PET and 18-month cognitive decline: a multicenter study. *Neurology*. 2012;79:1636-1644.
28. Doraiswamy PM, Sperling RA, Johnson K, et al. Florbetapir F 18 amyloid PET and 36-month cognitive decline: a prospective multicenter study. *Mol Psychiatry*. 2014;19:1044-1051. doi:10.1038/mp.2014.9.
29. Landau SM, Mintun MA, Joshi AD, et al. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. 2012;72(4):578-586. doi:10.1002/ana.23650.
30. Cronk BB, Johnson DK, Burns JM. Body mass index and cognitive decline in Mild Cognitive Impairment. 2010;24(2):126-130. doi:10.1097/WAD.0b013e3181a6bf3f.
31. Irizarry MC, Raman R, Schwarzschild MA, et al. Plasma urate and progression of

- Mild Cognitive Impairment. *Neurodegener Dis.* 2009;6:23-28. doi:10.1159/000170883.
32. Kennedy RE, Schneider LS, Cutter GR. Biomarker positive and negative subjects in the ADNI cohort: clinical characterization. *Curr Alzheimer Res.* 2012;9(10):1135-1141.
 33. Kennedy RE, Cutter GR, Schneider LS. Effect of APOE genotype status on targeted clinical trials outcomes and efficiency in dementia and MCI due to AD. 2014;10(3):349-359. doi:10.1016/j.jalz.2013.03.003.
 34. Kennedy RE, Cutter GR, Wang G, Schneider LS. Post hoc analyses of ApoE genotype-defined subgroups in clinical trials. 2016;50:1205-1215. doi:10.3233/JAD-150847.
 35. Mackin RS, Insel P, Tosun D, et al. The effect of subsyndromal symptoms of depression and white matter lesions on disability for individuals with Mild Cognitive Impairment. *Am J Geriatr Psychiatry.* 2013;21(9):906-914. doi:10.1016/j.jagp.2013.01.021.
 36. McGough EL, Cochrane BB, Pike KC, Logsdon RG, McCurry SM, Teri L. Dimensions of physical frailty and cognitive function in older adults with amnesic Mild Cognitive Impairment. *Ann Phys Rehabil Med.* 2013;56(5):329-341. doi:10.1016/j.rehab.2013.02.005.
 37. Portelius E, Zetterberg H, Skillbäck T, et al. Cerebrospinal fluid neurogranin: relation to cognition and neurodegeneration in Alzheimer's Disease. *Brain.* 2015;138:3373-3385. doi:10.1093/brain/awv267.
 38. Rozzini L, Chilovi BV, Bertolotti E, et al. Angiotensin converting enzyme (ACE) inhibitors modulate the rate of progression of amnesic Mild Cognitive Impairment. 2006;21:550-555. doi:10.1002/gps.1523.
 39. Rozzini L, Chilovi BV, Bertolotti E, et al. Mild parkinsonian signs and psycho-behavioral symptoms in subjects with Mild Cognitive Impairment. 2008;20(1):86-95. doi:10.1017/S1041610207006163.
 40. Schneider LS, Insel PS, Weiner MW. Treatment with Cholinesterase Inhibitors and Memantine of patients in the Alzheimer's Disease Neuroimaging Initiative. *Arch Neurol.* 2011;68(1):58-66. doi:10.1001/archneurol.2010.343.
 41. Toledo JB, Korff A, Shaw LM, Trojanowski JQ, Zhang J. Low levels of cerebrospinal fluid complement 3 and factor H predict faster cognitive decline in Mild Cognitive Impairment. *Alzheimers Res Ther.* 2014;6:36.
 42. Whitehair DC, Sherzai A, Emond J, et al. Influence of Apolipoprotein E e4 on rates of cognitive and functional decline in Mild Cognitive Impairment. *Alzheimers Dement.* 2010;6(5):412-419. doi:10.1016/j.jalz.2009.12.003.
 43. Bettermann K, Arnold AM, Williamson J, et al. Statins, risk of dementia and cognitive function: secondary analysis of the Ginkgo Evaluation of Memory Study (GEMS). 2012;21(6):436-444. doi:10.1016/j.jstrokecerebrovasdis.2010.11.002.
 44. Perneczky R, Pohl C, Sorg C, et al. Complex activities of daily living in Mild Cognitive Impairment: conceptual and diagnostic issues. *Age Ageing.* 2006;35:240-245. doi:10.1093/ageing/afj054.
 45. Lo RY, Hubbard AE, Shaw LM, et al. Longitudinal change of biomarkers in cognitive decline. *Arch Neurol.* 2011;68(10):1257-1266. doi:10.1001/archneurol.2011.123.
 46. Ye BS, Lee WW, Ham JH, Lee JJ, Lee PH, Sohn YH. Does serum uric acid act as

- a modulator of cerebrospinal fluid Alzheimer's Disease biomarker related cognitive decline? *Eur J Neurol*. 2016;23:948-957. doi:10.1111/ene.12969.
47. Cardinali DP, Vigo DE, Olivar N, Vidal MF, Furio AM, Brusco LI. Therapeutic application of melatonin in Mild Cognitive Impairment. *Am J Neurodegener Dis*. 2012;1(3):280-291.
 48. Furio AM, Brusco LI, Cardinali DP. Possible therapeutic value of melatonin in Mild Cognitive Impairment: a retrospective study. 2007;43:404-409. doi:10.1111/j.1600-079X.2007.00491.x.
 49. De Beaumont L, Pelleieux S, Lamarre-Théroux L, Dea D, Poirier J. Butyrylcholinesterase K and Apolipoprotein E-e4 reduce the age of onset of Alzheimer's Disease, accelerate cognitive decline, and modulate Donepezil response in Mild Cognitively Impaired subjects. *J Alzheimer's Dis*. 2016;54(3):913-922. doi:10.3233/JAD-160373.
 50. Buschert VC, Friese U, Teipel SJ, et al. Effects of a newly developed cognitive intervention in Amnesic Mild Cognitive Impairment and mild Alzheimer's Disease: a pilot study. 2011;25:679-694. doi:10.3233/JAD-2011-100999.
 51. Buschert VC, Giegling I, Teipel SJ, et al. Long-term observation of a multicomponent cognitive intervention in Mild Cognitive Impairment. *J Clin Psychiatry*. 2012;73(12):e1492-e1498. doi:10.4088/JCP.11m07270.
 52. Chiu C, Su K, Cheng T, et al. The effects of omega-3 fatty acids monotherapy in Alzheimer's Disease and Mild Cognitive Impairment: a preliminary randomized double-blind placebo-controlled study. 2008;32:1538-1544. doi:10.1016/j.pnpbp.2008.05.015.
 53. Dubois B, Zaim M, Touchon J, et al. Effect of six months of treatment with V0191 in patients with suspected prodromal Alzheimer's Disease. 2012;29:527-535. doi:10.3233/JAD-2012-111370.
 54. Förster S, Buschert VC, Buchholz H-G, et al. Effects of a 6-month cognitive intervention program on brain metabolism in amnesic Mild Cognitive Impairment and mild Alzheimer's Disease. *J Alzheimers Dis*. 2011;26:337-348. doi:10.3233/JAD-2011-0025.
 55. Kile S, Au W, Parise C, et al. IVIG treatment of Mild Cognitive Impairment due to Alzheimer's Disease: a randomised double-blinded exploratory study of the effect on brain atrophy, cognition and conversion to dementia. *J Neurol Neurosurg Psychiatry*. 2015;0:1-7. doi:jnnp-2015-311486.
 56. Lin C-H, Chen P-K, Chang Y-C, et al. Benzoate, a d-amino acid oxidase inhibitor, for the treatment of early-phase Alzheimer Disease: a randomized, double-blind, placebo-controlled trial. *Biol Psychiatry*. 2014;75:678-685. doi:10.1016/j.biopsych.2013.08.010.
 57. Luchsinger JA, Perez T, Chang H, et al. Metformin in Amnesic Mild Cognitive Impairment: results of a pilot randomized placebo controlled clinical trial. 2016;51:501-514. doi:10.3233/JAD-150493.
 58. Miao Y-C, Tian J-Z, Shi J, Mao M. Effects of Chinese medicine for tonifying the kidney and resolving phlegm and blood stasis in treating patients with amnesic Mild Cognitive Impairment: A randomized double-blind and parallel-controlled trial. *J Chinese Integr Med*. 2012;10(4):390-397. doi: 10.3736/jcim20120407.
 59. Petersen RC, Thomas RG, Grundman M, et al. Vitamin E and Donepezil for the treatment of Mild Cognitive Impairment. *NEJM*. 2005;352(23):2379-2388.

- doi:10.1056/NEJMoa1516783.
60. De Gobbi Porto FH, Coutinho AMN, De Sa Pinto AL, et al. Effects of aerobic training on cognition and brain glucose metabolism in subjects with Mild Cognitive Impairment. *J Alzheimer's Dis*. 2015;46:747-760. doi:10.3233/JAD-150033.
 61. Reuter I, Mehnert S, Sammer G, Oechsner M, Engelhardt M. Efficacy of a multimodal cognitive rehabilitation including psychomotor and endurance training in Parkinsons Disease. *J Aging Res*. 2012;2012:1-15. doi:10.1155/2012/235765.
 62. Singh MAF, Gates N, Saigal N, et al. The Study of Mental and Resistance Training (SMART) Study-resistance training and/or cognitive training in Mild Cognitive Impairment: a randomized, double-blind, double-sham controlled trial. *J Am Med Dir Assoc*. 2014;15:873-880. doi:10.1016/j.jamda.2014.09.010.
 63. Snitz BE, O'Meara ES, Carlson MC, Arnold AM, Ives DG. Ginkgo biloba for preventing cognitive decline in older adults: a randomized trial. *JAMA*. 2009;302(24):2663-2670. doi:10.1001/jama.2009.1913.
 64. Suzuki T, Shimada H, Makizako H, et al. A randomized controlled trial of multicomponent exercise in older adults with Mild Cognitive Impairment. *PLoS One*. 2013;8(4):e61483. doi:10.1371/journal.pone.0061483.
 65. Thal LJ, Ferris SH, Kirby L, et al. A randomized, double-blind, study of Rofecoxib in patients with Mild Cognitive Impairment. 2005;30:1204-1215. doi:10.1038/sj.npp.1300690.
 66. Caroli A, Prestia A, Wade S, et al. Alzheimer's Disease biomarkers as outcome measures for clinical trials in MCI. *Alzheimer Dis Assoc Disor*. 2015;29(2):101-109. doi:10.1097/WAD.0000000000000071.
 67. Grill JD, Di L, Lu PH, et al. Estimating sample sizes for pre-dementia Alzheimer's trials based on the Alzheimer's Disease Neuroimaging Initiative. *Neurobiol Aging*. 2013;34(1):62-72. doi:10.1016/j.neurobiolaging.2012.03.006.
 68. Ho AJ, Hua X, Lee S, et al. Comparing 3 T and 1.5 T MRI for tracking AD progression with tensor-based morphometry. 2010;l(4):499-514. doi:10.1002/hbm.20882.
 69. Hua X, Lee S, Yanovsky I, et al. Optimizing power to track brain degeneration in Alzheimer's Disease and Mild Cognitive Impairment with tensor-based morphometry: an ADNI study of 515 subjects. *Neuroimage*. 2009;48(4):668-681. doi:10.1016/j.neuroimage.2009.07.011.
 70. Mohs RC, Knopman D, Petersen RC, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. *Alzheimer Dis Assoc Disor*. 1997;11:S13-S21.
 71. Hobart J, Posner H, Aisen P, et al. The ADAS-Cog's performance as a measure-lessons from the ADNI study: Part 3 - do the scale modifications add value? In: *Poster Presentation*. 2009. doi:10.1017/CBO9781107415324.004.
 72. Ferris SH. Cognitive outcome measures. *Alzheimer Dis Assoc Disord*. 1999;13:S140-S142.
 73. Ferris SH. General measures of cognition. *Int Psychogeriatr*. 2003;15:215-217.
 74. Dichgans M, Markus HS, Salloway S, et al. Donepezil in patients with subcortical vascular cognitive impairment: a randomised double-blind trial in CADASIL. *Lancet Neurol*. 2008;7:310-318. doi:10.1016/S1474-4422(08)70046-2.

75. Lowe DA, Balsis S, Benge JF, Doody RS. Adding delayed recall to the ADAS-Cog improves measurement precision in mild Alzheimer's Disease: implications for predicting Instrumental Activities of Daily Living. *Psychol Assess.* 2015;27(4):1234-1240. doi:10.1037/pas0000133.
76. Labos E, Soderlund M, Trojanoski S, Seinhart D. Presentation of ADAS-Cog test with delayed recall and semantic encoding during the learning phase: results in normal, MCI and dementia population. *Alzheimer's Dement.* 2011;7(4):S250-S251.
77. Carusone SC, Goldsmith CH, Smieja M, Loeb M. Summary measures were a useful alternative for analyzing therapeutic clinical trial data. *J Clin Epidemiol.* 2006;59:387-392. doi:10.1016/j.jclinepi.2005.05.009.
78. Wouters H, van Gool WA, Schmand B, Lindeboom R. Revising the ADAS-Cog for a more accurate assessment of cognitive impairment. *Alzheimer Dis.* 2008;22(3):236-244.
79. O'Halloran J, Kemp A, Salmon D, Tariot P, Schneider L. Psychometric comparison of standard and computerized administration of the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog). *Curr Alzheimer Res.* 2011;8:323-328.
80. Inoue M, Jimbo D, Taniguchi M, Urakami K. Touch Panel-type Dementia Assessment Scale: a new computer-based rating scale for Alzheimer's Disease. *Psychogeriatrics.* 2011;11:28-33. doi:10.1111/j.1479-8301.2010.00345.x.
81. Wouters H, Van Campen J, Appels B, et al. Does adaptive cognitive testing combine efficiency with precision? Prospective findings. *J Alzheimer's Dis.* 2011;25:595-603. doi:10.3233/JAD-2011-101743.
82. Ihl R, Ferris S, Robert P, Winblad B, Gauthier S, Tennigkeit F. Detecting treatment effects with combinations of the ADAS-cog items in patients with mild and moderate Alzheimer's Disease. *Int J Geriatr Psychiatry.* 2012;27:15-21. doi:10.1002/gps.2679.
83. Balsis S, Unger A a., Benge JF, Geraci L, Doody RS. Gaining precision on the Alzheimer's Disease Assessment Scale-Cognitive: a comparison of item response theory-based scores and total scores. *Alzheimer's Dement.* 2012;8(4):288-294. doi:10.1016/j.jalz.2011.05.2409.
84. Verma N, Beretvas SN, Pascual B, Masdeu JC, Markey MK. New scoring methodology improves the sensitivity of the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) in clinical trials. *Alzheimers Res Ther.* 2015;7(64):1-17. doi:10.1186/s13195-015-0151-0.
85. Wessels A, Siemers E, Andersen S, et al. A combined measure of cognition and function for clinical trials: The integrated Alzheimer's Disease Rating Scale (iADRS). *J Prev Alzheimers Dis.* 2015;2(4):227-241. doi:10.1007/978-1-4614-5915-6.
86. Huang Y, Ito K, Billing CB, Anziano RJ. Development of a straightforward and sensitive scale for MCI and early AD clinical trials. *Alzheimer's Dement.* 2015;11(4):404-414. doi:10.1016/j.jalz.2014.03.008.
87. Grochowalski JH, Liu Y, Siedlecki KL. Examining the reliability of ADAS-Cog change scores. *Aging, Neuropsychol Cogn.* 2016;23(5).
88. Wang J, Logovinsky V, Hendrix SB, et al. ADCOMS: a composite clinical outcome for prodromal Alzheimer's Disease trials. *J Neurol Neurosurg Psychiatry.* 2016;0:1-7. doi:10.1136/jnnp-2015-312383.

Chapter 4

4 Research Question and Objectives

The purpose of Chapter 4 is to present the research question and objectives of this thesis.

4.1 Research Question

Does adding assessments of motor function to the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog 11) improve responsiveness among older adults with pre-dementia syndromes, wherein responsiveness is contextualized by the type of change being assessed?

We hypothesized that adding assessments of motor function to the ADAS-Cog 11 using a pooled index approach would improve responsiveness in a sample of older adults with Normal Cognition (NC), Subjective Cognitive Impairment (SCI), and Mild Cognitive Impairment (MCI).

4.2 Objectives

4.2.1 Objective 1

Use a pooled index approach to develop an outcome measure that can be backwards compatible to the ADAS-Cog 11 and includes measures of quantitative gait and dual-task gait cost.

4.2.2 Objective 2

Compare the responsiveness of the ADAS-Cog 11* and the novel outcome measure to group-level between-person differences in stage of pre-dementia disease progression at one point in time (baseline discrimination between groups of subjects classified as having NC, SCI, MCI).

4.2.3 Objective 3

Compare the responsiveness of the ADAS-Cog 11*, the novel outcome measure, and the ADAS-Cog combined, using a pooled index approach, with each individual component of the novel outcome measure to group-level within-person observed change over time in subjects with pre-dementia levels of impairment (disease progression or measured change over the course of a prospective cohort study).

4.3 Conclusion

The next chapter provides an overview of the rationale, methods, and results pertaining to the research question and objectives.

*Note: Secondary data analysis was used to achieve these objectives. Due to limitations in the availability of data, not all necessary assessments of older adults with pre-dementia syndromes were present in a single database. To obtain a preliminary answer to the research question, a statistical model was developed and then used to estimate ADAS-Cog 11 scores in a database with quantitative motor function assessments. Details about this “ADAS-Cog-Proxy” model are summarized in Chapter 5 (Section 5.2.2.1), and described in detail in Chapter 6 (Section 6.1.1).

Chapter 5

5 Integrated Article: Cognition and Motor Function: The Gait and Brain Pooled Index

Chapter 5 includes a version of a manuscript that will be submitted to the Alzheimer’s Disease Neuroimaging Initiative (ADNI) for review, and then pending approval by ADNI, to the *Journals of Gerontology: Medical Sciences* for publication. Due to the nature of an integrated manuscript thesis format, there is some overlap between the information presented in Chapter 5 and the rest of the thesis. Also, in the interest of clarity the ADAS-Cog 11 is simply referred to as the ADAS-Cog throughout Chapter 5. Please see Chapter 6 for more detailed methods and results, which go beyond what may be submitted for a peer reviewed publication.

In accordance with the ADNI data use agreement, on the by-line of the submitted manuscript, after the named authors, the phrase “for the Alzheimer’s Disease Neuroimaging Initiative*” will be included, with the asterisk referring to the following statement: “*Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf”.

5.1 Introduction

The Alzheimer’s Disease (AD) Assessment Scale–Cognitive Subscale (ADAS-Cog) was developed in 1984 for the purpose of assessing cognitive dysfunction in AD.¹ Since then, the ADAS-Cog became widely adopted for use in studies of AD and related disorders, and is now considered the ‘gold standard’ for assessing treatment efficacy in clinical trials of antimentia medications. However, two shifts in the field since its development have called into question the continued use of the original ADAS-Cog. The first shift has been

towards studying pre-dementia syndromes, such as Mild Cognitive Impairment (MCI), and testing interventions aimed at slowing or preventing progression to dementia rather than intervening at the dementia stage. The second shift includes postulating motor function as an important component of dementia and pre-dementia syndromes. These shifts have elicited the need for an outcome measure that reflects current research focus, incorporates all important disease components, and is more responsive than the ADAS-Cog to clinically important changes in pre-dementia syndromes.

In short, responsiveness is a form of validity defined as the ability to accurately detect change.²⁻⁵ Change can be contextualized using three aspects: group versus individual level of measurement, between-person versus within-person comparison, and the type of change one is interested in detecting.⁵ The responsiveness of any outcome measure is population and context specific.^{3,5}

While the ADAS-Cog has demonstrated responsiveness to multiple types of change in dementia populations, concerns have been raised about its responsiveness at pre-dementia stages where changes are subtler in magnitude and slower in rate of progression. Several modifications have improved the responsiveness of the ADAS-Cog in pre-dementia syndromes. These include alternative scoring applied to the original summation of errors made across ADAS-Cog tasks that gives a final score from 0 to 70, removing tasks with ceiling effects, and the addition of higher order assessments of delayed word recall, executive function, or the ability to perform activities required for daily independent living.⁶⁻¹⁴ The advantage of modified measures that are backwards compatible with the original ADAS-Cog is that they maintain consistency with previous studies and do not become a limiting factor if one wants to compare novel study findings with the large literature base that has used the ADAS-Cog. To ensure backwards compatibility, the ADAS-Cog must be administered in its original form so that raw scores may be obtained regardless of any modifications made thereafter.

The emergence of motor function decline as a potential biomarker for dementia and pre-dementia syndromes opens a unique opportunity for precise objective motor function tests to help assess severity or stages of disease impairment not captured by traditional cognitive tests.^{15,16} Poor performance on motor function tests has been associated with an

increased risk of disease progression,¹⁷⁻²² and combined cognitive and motor function impairments have been associated with a greater risk of further cognitive decline and conversion to dementia than either component alone.^{23,24} The mechanistic rationale for motor function being an integral component of dementia and pre-dementia syndromes is that neuropathology which contributes to cognitive impairment traditionally associated with dementia may also give rise to motor impairment.^{15,25,26} Brain regions hypothesized to underlie simultaneous decline in cognition and motor function, such as gait control, includes the frontal and temporal lobes, and frontal-hippocampal and thalamic-striatal circuits.^{15,25,26} Quantitative gait parameters, such as velocity or variability in step time, have been associated with concurrent levels of global and domain specific cognitive ability, have demonstrated the ability to discriminate between subtypes of MCI, and have shown responsiveness to changes in cognition over time.^{17,19,23,27-30} Changes in gait parameters between when a participant is asked to walk as they usually would and while performing a cognitive task, termed dual-task cost (DTC), have also been associated with cognitive abilities and pre-dementia syndromes.^{28,31} Importantly, the ability to maintain gait control while walking and thinking (low DTC) underlies functionality, defined as the ability to perform daily activities required for independent living, such as cooking and cleaning.³²

A review of the literature failed to reveal an outcome measure developed for any population that includes the addition of single-task motor assessments or DTC to the ADAS-Cog. We hypothesized that adding assessments of motor function to the ADAS-Cog would improve responsiveness among older adults with pre-dementia syndromes. Due to the lack of a database with both the ADAS-Cog and quantitative motor assessments, we developed a statistical model that uses alternative cognitive outcome measures to approximate ADAS-Cog scores (ADAS-Cog-Proxy). This ADAS-Cog-Proxy model was applied in a database that includes the necessary alternative cognitive measures and quantitative gait assessments. Our objectives were: 1) use a pooled index (PI) approach^{33,34} to develop an outcome measure that can be backwards compatible to the ADAS-Cog and includes quantitative gait and DTC assessments, 2) compare the responsiveness of the ADAS-Cog-Proxy and the novel PI to group-level between-person differences in stage of disease progression at one point in time (baseline discrimination),

and 3) compare the responsiveness of the ADAS-Cog-Proxy, the novel PI, and the ADAS-Cog-Proxy combined using a PI approach with each individual component of the novel PI, to group-level within-person measured change over time in a pre-dementia sample (disease progression).

5.2 Methods

5.2.1 Study Population

The Gait and Brain Study (GABS) is an ongoing prospective cohort study based in London, Ontario aimed at assessing how changes in gait may precede dementia and falls (Research Ethics Board approval number 17200). Participant recruitment began in 2007 from Geriatrics and Memory clinics at hospitals affiliated with Western University. Inclusion criteria were 65 to 85 years old, able to walk 10 meters without assistance, and absence of dementia. Exclusion criteria were lack of English proficiency, Parkinsonism or other neurological disorder affecting motor function (e.g. stroke), musculoskeletal disorders or joint replacements that affect gait performance (clinician assessed), osteoarthritis affecting lower limbs, use of psychotropics that can influence motor performance (e.g. benzodiazepines), and major depression. Eligible participants were divided into three diagnostic categories based on performance in cognitive testing and clinical evaluation. Normal Cognition (NC) criteria included normal age-, sex-, and education-adjusted scores on the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment. Subjective Cognitive Impairment (SCI) criteria were the same as that for NC, except patients reported persistent decline in cognition that was not explainable by an acute event, and answered yes to both, “Do you feel like your memory or thinking is becoming worse?” and “Does this concern you?”. MCI criteria included 1) a score of 0.5 on the Clinical Dementia Rating (CDR) Scale, 2) subjective cognitive complaints, 3) measured cognitive impairment in memory, executive function, attention, and/or language domains 4) intact Lawton-Brody Activities of Daily Living, and 5) absence of dementia based on Diagnostic and Statistical Manual of Mental Disorders version IV-TR or V criteria.^{35,36} Additional information can be found at [clinicaltrials.gov](https://clinicaltrials.gov/study/NCT03020381), study identifier NCT03020381.

5.2.2 Measures

5.2.2.1 Cognition

ADAS-Cog-Proxy scores were estimated in GABS using a generalized additive model (GAM)³⁷⁻³⁹ developed in the Alzheimer's Disease Neuroimaging Initiative (ADNI), which contains the ADAS-Cog as well as several cognitive measures administered in the GABS.

The ADNI began in 2003 as a public-private study partnership with the primary goal of testing whether neuroimaging, biological, clinical, and neuropsychological assessments can be combined to measure progression from MCI to early AD (adni.loni.usc.edu). Dr. Michael W. Weiner is the Principal Investigator of ADNI, and study sites are located throughout North America. Frequently updated information on ADNI can be found at www.adni-info.org. ADNI data was downloaded on October 26, 2016.

The process from ADAS-Cog-Proxy development to estimation in GABS is outlined in Figure 7 as five key steps (grey boxes), which are briefly described below.

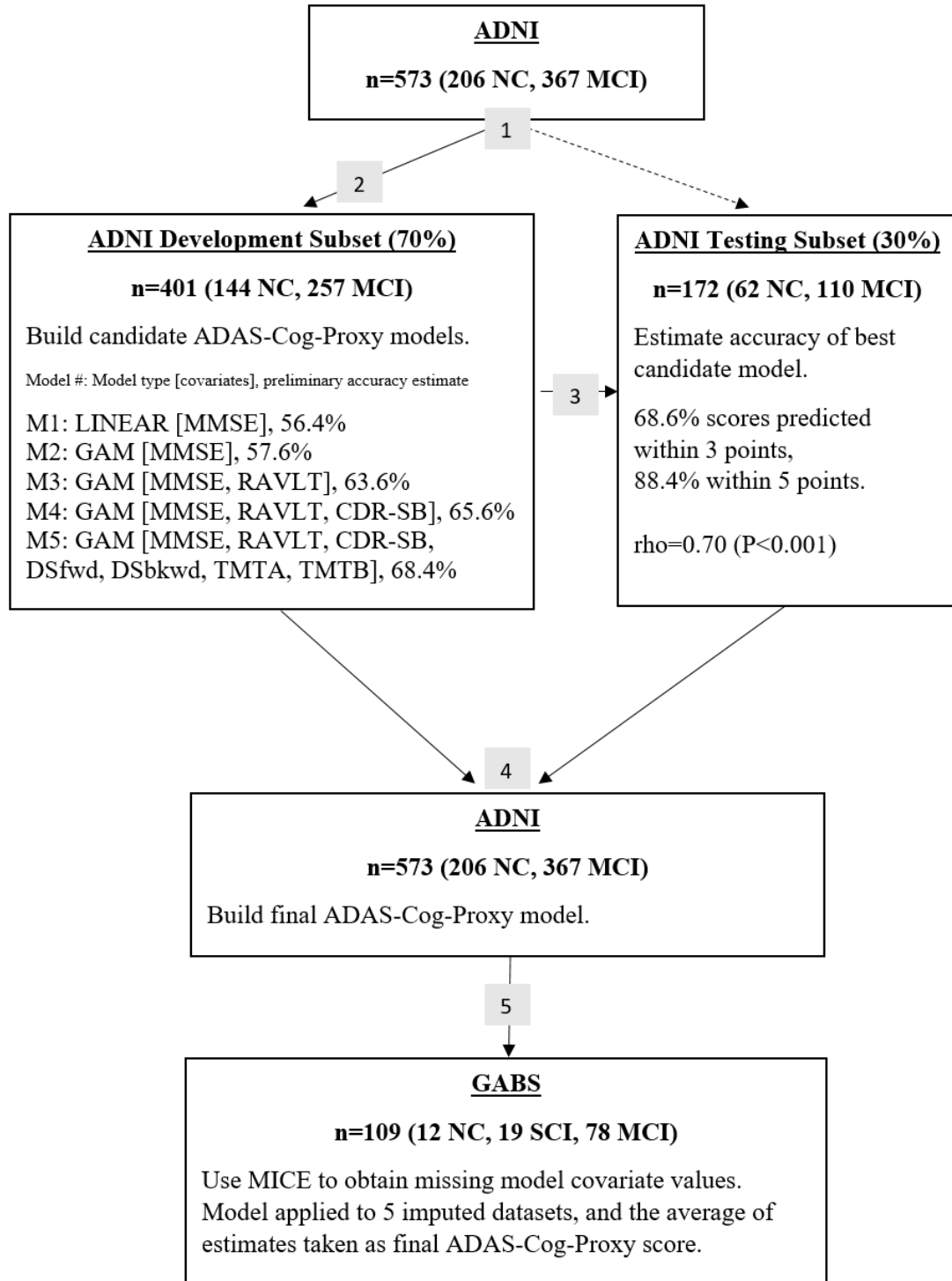


Figure 7 ADAS-Cog-Proxy model development and application.

Participants in the first of three ADNI phases (ADNI 1) with NC or MCI were divided into a 70% subset for model development and a 30% subset for testing model accuracy (Step 1). Five candidate models were constructed in the development subset (Step 2), and preliminary accuracy was assessed as the percentage (%) of participants for whom each candidate model predicted ADAS-Cog scores within three points of their observed ('true') score; three points is often considered a clinically relevant change.⁴⁰ Five ADNI participants were missing at least one covariate value for candidate Model 5 (M5) and were excluded solely from analyses that included M5. Diagnostics for all candidate models were assessed in the development subset (Appendix B, Figures B.1 to B.5).

The best candidate model was selected based on preliminary accuracy estimates in the development subset and on similarity of covariates to ADAS-Cog tasks such that they assessed cognitive domains covered by the ADAS-Cog (memory, language, praxis) with minimal coverage of additional areas. Accuracy of the best candidate model was estimated (Step 3) using 'new' ADNI participants in the testing subset as the percentage of participants who had scores predicted within three and five points of their true ADAS-Cog scores. Spearman's rank correlation (ρ) between predicted and observed scores was also calculated. The final GAM used for estimation of ADAS-Cog-Proxy scores in GABS was built on recombined development and testing subsets (Step 4).³⁸ In order to obtain ADAS-Cog-Proxy scores for all participants in GABS, Multivariate Imputation by Chained Equations (MICE) was used to impute missing GAM covariate values (Step 5).^{41,42} Predictor variable selection was guided by the suggestions of van Buuren et al.⁴¹ and included diagnostic, cognitive, functional, and motor assessments. Five imputed datasets were created using the imputation method of predictive mean matching. Imputed values were viewed to ensure plausibility, and imputation streams plotted to assess convergence (Appendix B, Figures B.6 to B.11). The ADAS-Cog-Proxy GAM was applied to each of the five imputed data sets, and the mean of the five estimated scores for each participant taken as their final ADAS-Cog-Proxy score. This process of imputing missing covariate values, applying the GAM to each completed dataset, and then averaging the five estimated ADAS-Cog-Proxy scores was repeated for 6, 12, 24, 36, and 48 month follow-up visits.

5.2.2.2 Motor function

Quantitative gait performance was assessed under four conditions using an electronic walkway system (GAITRite™). To avoid capturing acceleration and deceleration phases, start and end points of the walkway were marked one metre away from the ends of a 6-metre recording distance. The four testing conditions were one single-task condition where participants were asked to walk as they usually would, and three dual-task conditions (see below). Reliability of gait parameters under single and dual-task conditions have been described elsewhere.⁴³ The following spatio-temporal gait parameters were captured: velocity (cm/s), stride time (ms), step time (ms), stride length (cm), step length (cm), double support time (ms), swing time (ms), stride width (cm), stride velocity (cm/s), and cadence (steps/min). The Coefficient of Variation (CV=Standard Deviation (SD)/Mean*100) standardizes variability estimates to mean values, thus allowing direct comparison of variability across variables measured using different units. The CV was calculated for all gait parameters except velocity.

5.2.2.3 Motor-cognitive performance

The dual-task gait paradigm was used to capture motor-cognitive performance. The three dual-task gait conditions included walking as usual while: i) counting backwards from 100 by ones, ii) counting backwards from 100 by sevens, and iii) naming animals. Participants were not instructed to prioritize the cognitive or walking task. DTC (%) was calculated for three gait parameters under all three secondary task conditions, using the formula: [(single-task condition – dual-task condition)/single-task condition]*100.³¹ The three parameters of velocity, stride time, and stride time CV were selected based on literature supporting their importance in dementia and pre-dementia syndromes.^{16,28,31,43}

5.2.3 Data Analysis

5.2.3.1 Outcome measure development

Advantages of using a PI approach for our outcome measure are that it allows variables with different scoring ranges to be combined into a single summary score, and when

component variables have low pairwise correlations the SD of the derived score decreases as the number of variables increases.^{34,45} PI scores were obtained by first ensuring all variables were coded so that higher values indicate greater dysfunction, calculating Z-scores for each variable ($Z = (\text{observation} - \text{group mean}) / \text{SD}$), and then averaging those Z-scores.^{33,45} The statistical advantages of the PI diminish after six component variables, and are greatest when pairwise correlations are less than 0.2, and slightly less so up to 0.4.³⁴

Variable selection for our PI was thus guided by pairwise correlation coefficients and by theoretical considerations. For example, we aimed to include at least one variable from each of the categories of cognition, motor function, and motor-cognitive performance, by selecting at least one single-task and at least one DTC variable to combine with the ADAS-Cog-Proxy. To do this we first assessed pairwise correlations between the ADAS-Cog-Proxy and each of the single-task and DTC gait variables separately. Variables were retained when $|\rho| < 0.2$ or when $|\rho| = 0.2$ to 0.4 with evidence supporting that parameter's involvement in dementia or pre-dementia syndromes. Pairwise correlation coefficients were calculated for all retained single-task gait and DTC variables. In looking for at least one weakly intercorrelated pair, when numerical considerations were similar, we chose variables that had greater evidence from previous studies supporting their involvement in pre-dementia or dementia syndromes. When both numerical and theoretical considerations were similar, box plots were created to assess which of the contending individual gait or DTC parameters, if any, demonstrated a stepwise progression from NC to SCI to MCI diagnostic categories. Scatterplots were consulted to ensure low correlations were not the result of a strong non-linear relationship. Ease of assessment was also considered for both individual variables and the PI as a whole.

5.2.3.2 Baseline discrimination

Due to skewness and small sample sizes non-parametric tests were used to evaluate responsiveness to baseline discrimination. Kruskal-Wallis tests were used to assess whether the ADAS-Cog-Proxy and PI could detect a significant difference among the

diagnostic categories of NC, SCI, and MCI. Mann-Whitney U tests were used to assess all pairwise comparisons.

5.2.3.3 Change over time

Standardized Response Means (SRM=mean difference score/SD of difference score) were calculated for 6, 12, 24, 36, and 48 month follow-up periods for the PI, the ADAS-Cog-Proxy, and the ADAS-Cog-Proxy plus each individual component of the PI combined using a PI approach. Standardization was always performed with respect to the baseline distribution of participants present at the follow-up visit of interest. No distinction was made between diagnostic categories for SRM calculations.

All analyses were conducted with RStudio, version 1.0.136.⁴⁶

5.3 Results

5.3.1 Baseline Characteristics

Baseline characteristics for the 573 ADNI participants used to build the ADAS-Cog-Proxy GAM can be found in Table 9, and for the 109 participants in GABS in Table 10. One GABS participant with SCI did not have single-task gait recorded at baseline and was omitted from PI development and subsequent analyses. GABS participants who converted to dementia were included in analyses for time points prior to their dementia diagnosis. Two participants converted by six months of follow-up, one by 12 months, four by 24 months, and one by 36 months. A summary of the number of missing GAM covariates that were imputed using MICE can be found in Chapter 6 (Table 14).

Table 9 Alzheimer's Disease Neuroimaging Initiative Baseline Characteristics

| Characteristic | Overall (n=573) |
|--|------------------------------|
| Mean (SD) Minimum, Maximum Number of missing values (if applicable) <i>unless otherwise specified</i> | |
| Age (years) | 75.17 (6.56) 54.40, 89.60 |
| Education (years) | 15.84 (2.94) 6.00, 20.00 |

| | |
|---|-------------------------------------|
| Sex | |
| Female n (%) | 228 (40%) |
| Male | 345 |
| Alzheimer's Disease Assessment Scale-Cognitive Subscale | 9.51 (4.63) 0, 28 |
| Mini-Mental State Examination | 27.78 (1.84) 23, 30 |
| Rey Auditory Verbal Learning Test (3 trials) | 18.34 (5.64) 5, 38 |
| Clinical Dementia Rating Scale-Sum of Boxes | 1.02 (1.03) 0.00, 4.50 |
| Trail Making Test A | 41.41 (20.08) 17.00, 188.00 4 |
| Trail Making Test B | 114.8 (65.62) 34.0, 348.0 5 |
| Digit Span Forward Test | 6.64 (1.05) 4, 8 |
| Digit Span Backward Test | 4.71 (1.19) 0, 7 |

Table 10 Gait and Brain Study Baseline Characteristics

| Characteristic Mean (SD) Minimum, Maximum Number of missing values (if applicable) <i>unless otherwise specified</i> | Overall (n=109) | Normal Cognition (n=12) | Subjective Cognitive Impairment (n=19) | Mild Cognitive Impairment (n=78) |
|--|------------------------------|--|---|---|
| Age (years) | 74.22 (6.33) 63.00, 92.00 | 73.50 (4.58) 67.00, 82.00 | 70.00 (4.59) 65.00, 85.00 | 75.36 (6.52) 63.00, 92.00 |
| Education (years) | 13.85 (2.92) 6.00, 20.00 | 16.33 (3.06) 10.00, 20.00 | 14.42 (2.81) 10.00, 20.00 | 13.33 (2.74) 6.00, 20.00 |
| Sex | | | | |
| Female n (%) | 58 (53) | 7 (58) | 15 (79) | 36 (49) |
| Male | 51 | 5 | 4 | 42 |
| Medications (#) | 7.62 (4.52) 0, 21 | 6.42 (4.06) 2, 16 | 6.53 (5.26) 0, 21 | 8.06 (4.37) 0, 21 |
| Comorbidities (#) | 6.06 (2.85) 0, 13 | 4.33 (1.44) 2, 7 | 4.79 (2.02) 1, 8 | 6.64 (2.98) 0, 13 |
| Geriatric Depression Scale | 2.35 (2.14) 0, 10 22 | 1.60 (1.14) 0, 3 7 | 2.25 (1.89) 1, 5 15 | 2.40 (2.21) 0, 10 0 |

| | | | | |
|--|--------------------------------------|----------------------------------|-----------------------------------|----------------------------------|
| General Physical Activity Level | | | | |
| Vigorous: n (%) | 63 (58) | 6 (50) | 13 (68) | 44 (56) |
| Moderate: n (%) | 29 (27) | 5 (42) | 4 (21) | 20 (26) |
| Seldom: n (%) | 16 (25) | 1 (8) | 2 (11) | 13 (17) |
| Missing: n | 1 | | | 1 |
| Lawton-Brody Activities of Daily Living | 5.99 (0.11) 5, 6 22 | 6.00 (0.00) 6, 6 7 | 5.75 (0.50) 5, 6 15 | 6.00 (0.00) 6, 6 0 |
| Instrumental Activities of Daily Living | 7.69 (0.94) 2, 8 22 | 8.00 (0.00) 8, 8 7 | 7.75 (0.50) 7, 8 15 | 7.67 (0.99) 2, 8 0 |
| Basic Activities of Daily Living | 0.42 (0.97) 0, 5 22 | 0.80 (0.84) 0, 2 7 | 0.75 (0.96) 0, 2 15 | 0.38 (0.98) 0, 5 |
| ADAS-Cog-Proxy | 9.46 (2.34) 3, 16 | 7.59 (1.32) 4, 9 | 7.96 (1.93) 3, 12 | 10.11 (2.24) 5, 16 |
| Montreal Cognitive Assessment | 24.45 (3.82) 12, 30 | 27.25 (1.48) 24, 30 | 27.89 (2.45) 21, 30 | 23.18 (3.60) 12, 30 |
| Mini-Mental State Examination | 27.74 (2.52) 18, 30 | 28.83 (1.80) 24, 30 | 28.89 (1.45) 24, 30 | 27.29 (2.69) 18, 30 |
| Clinical Dementia Rating Scale | 0.99 (0.89) 0.0, 4.0 68 | 0.0 (0.0) 0.0,0.0 9 | | 1.07 (0.88) 0.0, 4.0 40 |
| Rey Auditory Verbal Learning Test (3 trials) | 17.20 (5.35) 8.0, 33.0 29 | 23.40 (5.18) 19.0, 32.0 7 | 24.75 (6.65) 17.0, 33.0 15 | 16.34 (4.71) 8.0, 28.0 7 |
| Gait Velocity (cm/s) | 108.40 21.27 57.27, 165.2 1 | 124.80 15.78 99.65, 155.80 | 114.10 17.59 82.17, 141.00 | 104.60 21.47 57.27, 165.20 |
| Stride Time (s) | 1.14 (0.10) 0.93, 1.41 1 | 1.11 (0.08) 0.95, 1.20 | 1.10 (0.08) 0.97, 1.26 1 | 1.16 (0.10) 0.93, 1.41 |
| Stride Time Coefficient of Variation (CV) (%) | 2.47 (1.48) 0.62, 9.73 1 | 2.08 (0.76) 1.14, 4.04 | 2.49 (2.02) 1.16, 9.73 1 | 2.53 (1.43) 0.62, 7.89 |
| Dual-Task Gait Velocity Cost with Counting (%) | 5.51 (10.68) -16.04, 34.61 1 | 3.10 (11.52) -8.16, 34.61 | 2.58 (5.48) -11.05, 10.82 1 | 6.55 (11.35) -16.04, 31.12 |

| | | | | |
|---|--|---------------------------------------|---|---|
| Dual-Task Stride Time Cost with Serial Sevens (%) | -16.93 (18.42) -75.93, 6.30 3 | -24.06 (29.08) -75.93, 3.74 | -8.23 (9.87) -38.54, 2.74 1 | -17.86 (17.37) -69.50, 6.30 2 |
| Dual-Task Stride Time CV Cost with Naming Animals (%) | -133.40 (270.66) -1382.00, 77.58 1 | -214.80 (416.11) -1382.0, 63.87 | -44.54 (80.73) -240.3, 53.55 1 | -141.30 (269.59) -1200.00, 77.58 |

5.3.2 ADAS-Cog Proxy Model

Covariates for the GAM selected to estimate ADAS-Cog-Proxy scores included the sum of the first three trials of the Rey Auditory Verbal Learning Test (RAVLT), the MMSE, and the CDR-Sum of Boxes (CDR-SB) score. This ADAS-Cog-Proxy model estimated 69% of participant scores within three points and 88% of participant scores within five points of their observed ('true') ADAS-Cog score in the testing subset of ADNI (Figure 7). Spearman's rank correlation coefficient between predicted and observed ADAS-Cog scores was 0.70 ($P < 0.001$). Baseline ADAS-Cog-Proxy scores in GABS are included in Table 10.

5.3.3 Gait and Brain Pooled Index

Variables selected for inclusion in the PI include the ADAS-Cog-Proxy, gait velocity multiplied by negative one, and DTC for gait velocity with the secondary task of counting backwards from 100 by ones. Pairwise correlation coefficients ranged in magnitude from 0.27 to 0.32 (Chapter 6, Table 20).

5.3.4 Baseline Discrimination

Both the ADAS-Cog-Proxy and the PI showed an overall statistically significant difference in mean ranks across the three diagnostic categories (ADAS-Cog-Proxy: Kruskal-Wallis $H(2)$ value=24.13; PI: $H(2)$ =22.36, both $P < 0.001$). Statistically significant pairwise comparisons were found for SCI versus MCI (ADAS-Cog-Proxy: Mann-Whitney U test statistic=331, $P=0.0002$; PI: $U=348$, $P=0.0009$) and NC versus MCI

(ADAS-Cog-Proxy: $U=153$, $P=0.0002$; PI: $U=148$, $P=0.0001$), but not NC versus SCI diagnostic categories (ADAS-Cog-Proxy: $U=93$, $P=0.41$; PI: $U=75$, $P=0.17$).

5.3.5 Change Over Time

Adding only gait velocity to the ADAS-Cog-Proxy using a PI approach always increased responsiveness to decline (less negative or more positive SRM), while adding only DTC to the ADAS-Cog-Proxy showed mixed results (Table 11). The full PI had a larger SRM than the ADAS-Cog-Proxy for 6 month (SRM: ADAS-Cog-Proxy=0.14, PI=0.23) and 48 month (SRM: ADAS-Cog-Proxy=0.60, PI=0.65), but not 36 month (SRM: ADAS-Cog-Proxy=0.23, PI=0.18) follow-up periods. For 12 and 24 month follow-up periods the ADAS-Cog-Proxy detected overall improvement (SRM: 12 month=-0.08, 24 month=-0.24), while the full PI detected almost no change (SRM: 12 month=0.04, 24 month=0.01).

Table 11 Standardized Response Means: Responsiveness to group-level within-person measured change over time

| n | Time | ADASp | ADASp+GV | ADASp+DTC | ADASp+GV+DTC (Full PI) |
|----|------|-------|----------|-----------|------------------------|
| 86 | 6m | 0.14 | 0.17 | 0.18 | 0.23 |
| 73 | 12m | -0.08 | -0.05 | 0.01 | 0.04 |
| 55 | 24m | -0.24 | -0.11 | -0.07 | 0.01 |
| 35 | 36m | 0.23 | 0.34 | 0.11 | 0.18 |
| 24 | 48m | 0.60 | 0.68 | 0.59 | 0.65 |

Notes: Comparisons of the magnitude of standardized response means should only be made across rows because due to the nature of using data from an ongoing cohort study, not all participants have had the chance to reach all follow-up visits, and participants who converted to dementia were only included in calculations before the point of conversion.

Legend: ADASp=Alzheimer's Disease Assessment Scale-Cognitive Subscale-Proxy, DTC=Dual Task Cost (Gait Velocity (GV) with secondary task of counting backwards by ones), PI=Pooled Index, m=months.

5.4 Discussion

A PI approach combining assessments of motor function, specifically gait velocity and DTC gait velocity, with an ADAS-Cog-Proxy cognitive measure demonstrated comparable responsiveness to baseline discrimination between pre-dementia diagnostic categories and generally comparable or increased responsiveness to measured change over time as compared to the ADAS-Cog-Proxy cognitive measure alone.

More specifically, both the PI and ADAS-Cog-Proxy detected statistically significant differences between NC and MCI, and SCI and MCI predementia diagnostic categories but not between NC and SCI. This latter finding may have been due to small sample sizes rather than an inability to distinguish between the two mildest stages of disease progression. For all but one follow-up period the PI demonstrated greater responsiveness than the ADAS-Cog-Proxy to measured decline over time; however, there were two follow-up periods where the ADAS-Cog-Proxy detected improvement while the PI detected worsening. This group-level improvement measured by the ADAS-Cog-Proxy may be capturing the fact that the cognitive trajectory from NC to dementia is not linear such that some of the participants with MCI or SCI may have reverted to more normal levels of cognition. Also, excluding participants after conversion to dementia removed the participants who are expected to have experienced the largest decline. Further research is needed to assess whether the PI is detecting a more realistic overall assessment of the change in functionality over time than the ADAS-Cog-Proxy, as both cognitive and motor function are important for everyday living.

The improvements in responsiveness to group-level within-person measured change over time that occurred by adding gait and DTC assessments to the ADAS-Cog-Proxy were made without adding tests of delayed recall or executive function which have previously been found to improve the responsiveness of the ADAS-Cog in pre-dementia syndromes.⁸⁻¹¹ These cognitive abilities are thought to be important in pre-dementia syndromes but are not included on the original ADAS-Cog.^{1,9,47} Our results align with research exploring motor function as a biomarker for cognitive impairment and pre-dementia syndromes.^{15,28}

Key advantages of using quantitative gait assessments for outcome measurement include language independence, non-invasive administration procedures, avoidance of ceiling effects across the disease spectrum, and when DTC is used to assess the impact of a secondary cognitive task each participant serves as their own control. Further advantages of gait velocity are that it can be easily measured using only a stop watch and defined walking distance.

In addition to our findings that gait parameter tests of motor function may be valuable additions to cognitive assessments for use in pre-dementia studies, the creation of the ADAS-Cog-Proxy may provide a framework when there is an appropriate research question but not all necessary variables present in a single available database. Using a predictive model to obtain estimates of a missing variable allows preliminary tests of hypotheses without the time and resources that would be required to collect new data.

Main limitations of our study include small sample sizes, missing data, and reliance on a 'proof of principle' approach as we were unable to use the original ADAS-Cog. Two ADAS-Cog-Proxy GAM covariates were collected one month prior to the ADAS-Cog administration, which may have contributed extra noise to the GAM development and led to an underestimate of accuracy. Furthermore, the inclusion of only two variables on top of the ADAS-Cog-Proxy in our PI did not take advantage of statistical advantages that may be gained by including additional lowly correlated yet informative variables, such as other gait parameters or cognitive tests. Restricting our PI to only gait velocity single and DTC with the ADAS-Cog-Proxy represents the trade-off in information value between practicality and measurement intensiveness. The derived units of the PI are also difficult to interpret and are not directly comparable to ADAS-Cog scores. Selection bias, such as towards highly educated participants, may also limit the generalizability of our results.

Future steps include re-creating the PI using the original ADAS-Cog, assessing responsiveness with new participants across all levels of disease severity from NC to dementia, and assessing responsiveness to treatment effects in pre-dementia populations.

5.5 References

1. Rosen WG, Mohs RC, Davis KL. A New Rating Scale for Alzheimer's Disease. *Am J Psychiatry*. 1984;141:1356-1364. doi:10.1176/ajp.141.11.1356.
2. Krishner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis*. 1985;38(1):27-36.
3. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness. *J Clin Epidemiol*. 2000;53(5):459-468. doi:10.1016/S0895-4356(99)00206-1.
4. De Bruin AF, Diederiks JPM, De Witte LP, Stevens FCJ, Philipsen H. Assessing the responsiveness of a functional status measure: The Sickness Impact Profile versus the SIP68. *J Clin Epidemiol*. 1997;50(5):529-540. doi:10.1016/S0895-4356(97)00047-4.
5. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol*. 2001;54(12):1204-1217. doi:10.1016/S0895-4356(01)00407-3.
6. Wouters H, van Gool WA, Schmand B, Lindeboom R. Revising the ADAS-Cog for a more accurate assessment of cognitive impairment. *Alzheimer Dis*. 2008;22(3):236-244.
7. Llano D, Laforet G, Devanarayan V. Derivation of a new ADAS-Cog composite using tree-based multivariate analysis: prediction of conversion from Mild Cognitive Impairment to Alzheimer Disease. *Alzheimer Dis Assoc Disord*. 2011;25(1):73-84. doi:10.1097/WAD.0b013e3181f5b8d8.
8. Raghavan N, Samtani MN, Farnum M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimer's Dement*. 2013;9(10):S21-S31. doi:10.1016/j.jalz.2012.05.2187.
9. Skinner J, Carvalho JO, Potter GG, Thames A, Zelinski E. The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): an expansion of the ADAS-Cog to improve responsiveness in MCI. *Brain Imaging Behav*. 2012;6(4):489-501. doi:10.1007/s11682-012-9166-3.
10. Lowe DA, Balsis S, Bengtson JF, Doody RS. Adding delayed recall to the ADAS-Cog improves measurement precision in mild Alzheimer's Disease: implications for predicting Instrumental Activities of Daily Living. *Psychol Assess*. 2015;27(4):1234-1240. doi:10.1037/pas0000133.
11. Sano M, Raman R, Emond J, et al. Adding delayed recall to the Alzheimer Disease Assessment Scale is useful in studies of Mild Cognitive Impairment but not Alzheimer Disease. *Alzheimer Dis Assoc Disord*. 2011;25(2):122-127.
12. Wessels A, Siemers E, Andersen S, et al. A combined measure of cognition and function for clinical trials: the integrated Alzheimer's Disease Rating Scale (iADRS). *J Prev Alzheimers Dis*. 2015;2(4):227-241. doi:10.1007/978-1-4614-5915-6.
13. Huang Y, Ito K, Billing CB, Anziano RJ. Development of a straightforward and sensitive scale for MCI and early AD clinical trials. *Alzheimer's Dement*. 2015;11(4):404-414. doi:10.1016/j.jalz.2014.03.008.
14. Wang J, Logovinsky V, Hendrix SB, et al. ADCOMS: a composite clinical outcome for prodromal Alzheimer's Disease trials. *J Neurol Neurosurg Psychiatry*. 2016;jnnp-2015-312383. doi:10.1136/jnnp-2015-312383.
15. Montero-Odasso M. Gait as a biomarker of cognitive impairment and dementia

- syndromes. Quo vadis? *Eur J Neurol*. 2016;23(3):437-438. doi:10.1111/ene.12908.
16. Montero-Odasso M, Verghese J, Beauchet O, Hausdorff JM. Gait and cognition: a complementary approach to understanding brain function and the risk of falling. *J Am Geriatr Soc*. 2012;60(11):2127-2136. doi:10.1111/j.1532-5415.2012.04209.x.
 17. Mielke MM, Roberts RO, Savica R, et al. Assessing the temporal relationship between cognition and gait: slow gait predicts cognitive decline in the Mayo Clinic Study of Aging. *Journals Gerontol Ser A Biol Sci Med Sci*. 2013;68(8):929-937. doi:10.1093/gerona/gls256.
 18. Deshpande N, Metter EJ, Bandinelli S, Guralnik J, Ferrucci L. Gait speed under varied challenges and cognitive decline in older persons: a prospective study. *Age Ageing*. 2009;38(5):509-514. doi:10.1093/ageing/afp093.
 19. Camicioli R, Howieson D, Oken B, Sexton G, Kaye J. Motor slowing precedes impairment in the oldest old. *Neurology*. 1998;50(5):1496-1498. doi:10.1212/WNL.50.5.1496.
 20. Buracchio T, Dodge HH, Howieson D, Wasserman D, Kaye J. The trajectory of gait speed preceding Mild Cognitive Impairment. *Arch Neurol*. 2010;67(8):980-986. doi:10.1001/archneurol.2010.159.
 21. Beauchet O, Annweiler C, Callisaya ML, et al. Poor gait performance and prediction of dementia: results from a meta-analysis. *J Am Med Dir Assoc*. 2016. doi:10.1016/j.jamda.2015.12.092.
 22. Kueper JK, Speechley M, Lingum N, Montero-Odasso M. Motor function and incident dementia: a systematic review and meta-analysis. *Age Ageing*. 2017:1-10. doi: 10.1093/ageing/afx084
 23. Montero-Odasso MM, Barnes B, Speechley M, et al. Disentangling Cognitive-Frailty: results from the Gait and Brain Study. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2016;71(11):1476-1482. doi:10.1093/gerona/glw044.
 24. Verghese J, Annweiler C, Ayers E, et al. Motoric Cognitive Risk Syndrome multicountry prevalence and dementia risk. *Neurology*. 2014;83:718-726.
 25. Montero-Odasso M, Hachinski V. Preludes to brain failure: executive dysfunction and gait disturbances. *Neurol Sci*. 2014;35(4):601-604. doi:10.1007/s10072-013-1613-4.
 26. Annweiler C, Beauchet O, Bartha R, et al. Motor cortex and gait in Mild Cognitive Impairment: a magnetic resonance spectroscopy and volumetric imaging study. *Brain*. 2013;136(3):859-871. doi:10.1093/brain/aws373.
 27. Beauchet O, Allali G, Montero-Odasso M, Sejdić E, Fantino B, Annweiler C. Motor phenotype of decline in cognitive performance among community-dwellers without dementia: population-based study and meta-analysis. *PLoS One*. 2014;9(6):e99318. doi:10.1371/journal.pone.0099318.
 28. Montero-Odasso M, Oteng-Amoako A, Speechley M, et al. The motor signature of Mild Cognitive Impairment: results from the Gait and Brain Study. *Journals Gerontol-Ser A Biol Sci Med Sci*. 2014;69(11):1415-1421. doi:10.1093/gerona/glu155.
 29. Verghese J, Robbins M, Holtzer R, et al. Gait dysfunction in Mild Cognitive Impairment Syndromes. *J Am Geriatr Soc*. 2008;56(7):1244-1251. doi:10.1111/j.1532-5415.2008.01758.x.
 30. Allali G, Annweiler C, Blumen H, et al. Gait phenotype from MCI to moderate dementia: results from the GOOD initiative. *Eur J Neurol*. 2016;23(3):527-541.

- doi:10.1002/aur.1474.Replication.
31. Montero-Odasso M, Muir SW, Speechley M. Dual-task complexity affects gait in people with Mild Cognitive Impairment: the interplay between gait variability, dual tasking, and risk of falls. *Arch Phys Med Rehabil.* 2012;93(2):293-299. doi:10.1016/j.apmr.2011.08.026.
 32. Hausdorff JM, Yogev G, Springer S, Simon ES, Giladi N. Walking is more like catching than tapping: gait in the elderly as a complex cognitive task. *Exp Brain Res.* 2005;164(4):541-548. doi:10.1007/s00221-005-2280-3.
 33. Smythe H, Helewa A, Goldsmith C. "Independent assessor" and "pooled index" as techniques for measuring treatment effects in rheumatoid arthritis. *J Rheumatol.* 1977;4(2):144-152.
 34. Goldsmith C, Smythe H, Helewa A. Interpretation and power of a pooled index. *J Rheumatol.* 1993;20(3):575-578.
 35. Petersen RC. Mild Cognitive Impairment. *N Engl J Med.* 2011;364:2227-2234. doi:10.1017/S1355617714000137.
 36. Dubois B, Albert ML. Amnesic MCI or prodromal Alzheimer's Disease? *Lancet Neurol.* 2004;3(4):246-248. doi:10.1016/S1474-4422(04)00710-0.
 37. Wood SN. *Generalized Additive Models: An Introduction with R.* Boca Raton: Taylor & Francis Group; 2006.
 38. Wood S. Package 'mgcv'. 2017. <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>
 39. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1-22.
 40. Schrag A, Schott JM. What is the clinically relevant change on the ADAS-Cog? *J Neurol Neurosurg Psychiatry.* 2012;83(2):171-173. doi:10.1136/jnnp-2011-300881.
 41. van Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate imputation by chained equations. *J Stat Softw.* 2011;45(3):1-67.
 42. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med.* 2011;30(4):377-399. doi:10.1002/sim.4067.
 43. Montero-Odasso M, Casas A, Hansen KT, et al. Quantitative gait analysis under dual-task in older people with Mild Cognitive Impairment: a reliability study. *J Neuroeng Rehabil.* 2009;6(1):35. doi:10.1186/1743-0003-6-35.
 44. Montero-Odasso M, Bergman H, Phillips NA, Wong CH, Sourial N, Chertkow H. Dual-tasking and gait in people with Mild Cognitive Impairment. The effect of working memory. *BMC Geriatr.* 2009;9(1):41. doi:10.1186/1471-2318-9-41.
 45. Speechley M, Forchuk C, Hoch J, Jensen E, Wagg J. Deriving a mental health outcome measure using the pooled index: an application to psychiatric consumer-survivors in different housing types. *Heal Serv Outcomes Res Methodol.* 2009;9(2):133-143. doi:10.1007/s10742-009-0044-4.
 46. R Studio Team. R Studio. 2016. <http://www.rstudio.com/>.
 47. Mohs RC, Knopman D, Petersen RC, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. *Alzheimer Dis Assoc Disor.* 1997;11:S13-S21.

Chapter 6

6 Detailed Methods

The purpose of Chapter 6 is to provide a more detailed description of methods and supplementary results that are not presented in Chapter 5. To minimize redundancy not all methods details and results presented in Chapter 5 also appear in Chapter 6, but some overlap was necessary to maintain comprehension. Chapter 6 is organized according to thesis objective, and information on the development of the Alzheimer’s Disease Assessment Scale-Cognitive Subscale-Proxy measure (ADAS-Cog-Proxy) is included under the first objective.

6.1 Objective 1

Objective 1 was completed in a three-step process. First, the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database was used to build an ADAS-Cog-Proxy statistical model which could be used to obtain estimates of Alzheimer Disease Assessment Scale-Cognitive Subscale (ADAS-Cog 11) scores in the Gait and Brain Study (GABS) database. This step was necessary as there is no database with both the ADAS-Cog 11 and quantitative motor assessments, and developing a proxy measure to use a ‘proof of principle’ approach to the objectives allows a preliminary test of hypotheses before investing time and resources in a new study that could collect all necessary measures together. Second, additional measures in GABS were selected to add to the ADAS-Cog-Proxy. Third, the selected measures and ADAS-Cog-Proxy were combined using a pooled index (PI) approach.

6.1.1 Step 1: Develop an Alzheimer’s Disease Assessment Scale – Cognitive Subscale Proxy

ADNI wave selection and data obtainment. ADNI contains three waves of participants: ADNI 1, ADNI Grand Opportunities, and ADNI 2. Of these, ADNI 1 has the largest overlap of available cognitive tests with GABS (Table 12), and therefore was selected to build the ADAS-Cog-Proxy.

Table 12 Cognitive Tests in Available Databases at Baseline

| Tests | GABS | ADNI1 |
|----------------------------|------|-------------------------------|
| ADAS-Cog 11 Total | | X |
| ADAS-Cog Items | | X |
| MMSE Total | X | X |
| MMSE Items | | X |
| MoCA | X | |
| MoCA Subscores | X | |
| CDR-SB | X | X |
| TMTA | X | X |
| TMTB | X | X |
| Digit Span Forward | X | X |
| Digit Span Backward | X | X |
| Letter Number Sequence | X | |
| RAVLT (sum of 3 trials) | X | X (via item-level data) |
| BNT | X | |
| FAB | X | |

Legend: X indicates the test listed in the leftmost column of the row is present in the database at the column head. GABS=Gait and Brain Study, ADNI 1=Alzheimer's Disease Neuroimaging Initiative wave 1, ADAS-Cog 11=Alzheimer's Disease Assessment Scale-Cognitive Subscale 11 item version, MMSE=Mini-Mental State Examination, MoCA=Montreal Cognitive Assessment, CDR-SB=Clinical Dementia Rating Scale-Sum of Boxes, TMTA= Trail Making Test Part A, TMTB=TMT Part B, RAVLT=Rey Auditory Verbal Learning Test (immediate recall), BNT=Boston Naming Test, FAB=Frontal Assessment Battery. Note that ADNI administered a modified version of BNT whereby only odd questions were used. This modified version was not directly comparable with the full BNT version administered in GABS, and item level data was not available in GABS to create an odd question only version.

Inclusion criteria for ADNI 1 was Hachinski score less than or equal to 4, aged 55 to 90 years old, stability of ADNI permitted medications, Geriatric Depression Scale less than 6, study partner with at least 10 hours of contact with the participant per week, visual and auditory acuity adequate for neuropsychological testing, good general health, unable to bear children, willing and able to complete three year imaging study including no medical contraindications to magnetic resonance imaging, education level of grade 6 or work history, fluent English or Spanish speaking ability, agrees to DNA for Apolipoprotein

(APOE) e4 allele testing and banking, agrees to blood and urine samples for biomarker testing, and not enrolled in other trials or studies. There were further inclusion criteria specific to diagnostic categories of Normal Control or Cognition (NC), Mild Cognitive Impairment (MCI), and Alzheimer’s Disease (AD) groups. Inclusion criteria for NC was no abnormal memory complaints, normal memory function scores on the Logical Memory II subscale, Mini-Mental State Exam (MMSE) score between 24 and 30, Clinical Dementia Rating (CDR) score of 0 and Memory Box score of 0, and no abnormal levels of cognitive function or activities of daily living. Inclusion criteria for MCI categorization was memory complaint by the participant or participant’s study partner, abnormal memory function score on the Logical Memory II subscale, MMSE between 24 and 30, CDR score of 0.5 with Memory Box score of at least 0.5, and insufficient cognitive and functional impairment to allow a diagnosis of AD. Inclusion criteria for AD will not be reviewed as those participants were excluded from analyses in this thesis.

The online ADNI database is divided into several data tables, which are defined by the types of variables they contain. Some overlap between data tables exists. The “Item Level Data (ADAS-Cog, ANART, MMSE, etc) [ADNI1]” and “Key ADNI tables merged into one table” data tables were downloaded from the ADNI website (<http://adni.loni.usc.edu/data-samples/access-data/>) on October 26, 2016. All cognitive test scores were treated as numeric variables. All cognitive tests of interest were administered at a single baseline visit except the MMSE and CDR-SB, which were administered one month earlier at the ADNI 1 screening visit.

Baseline observations for participants in ADNI 1 with NC or MCI diagnostic status were retained from the “Key ADNI tables merged into one table” data table.

Baseline data from the “Item Level Data (ADAS-Cog, ANART, MMSE, etc) [ADNI1]” data table was used to create a *three trial Rey Auditory Verbal Learning Test (RAVLT) score*. The RAVLT involves reading a list of 15 words to a participant and then asking the participant to recall as many words as possible immediately after the list is read (trial 1). The same list of words is read a second time, and the participant is given up to four more trials (trials 2 to 5) to recall as many of the words as possible. ADNI performed a total of five trials while GABS performed three. To create a compatible three trial

summary score in ADNI, the number of words recalled on the first three trials were treated as numerical variables (0=not recalled, 1=recalled for each of the 15 words on each of the three trials) and summed together to give a score from 0 to 45. Twelve participants without item level RAVLT data were excluded.

This RAVLT sum of 3 trials score as well as total test scores for the Digit Span Forward and Backward Tests, and Trail Making Test parts A and B (TMT A & B), which were also included in the “Item Level Data (ADAS-Cog, ANART, MMSE, etc) [ADNI1]” data table, were merged with the selected “Key ADNI tables merged into one table” observations using Roster Identification Number (RID) such that observations needed to be included in both data tables (matching RID) to be retained. This method of merging ensured that participants with AD from the “Item Level” data, where diagnostic information was not recorded, were excluded. After merging, four participants did not have TMT A scores and five did not have TMT B scores. These participants were only excluded from analyses which required TMT A or B scores (candidate ADAS-Cog-Proxy Model 5).

Data splitting. ADNI data was split into 70% development (n=401) and 30% testing (n=172) subsets using random sampling without replacement via the `sample.split` command from the R package `caTools` with seed value set at 100.¹

Candidate model building and selection in development subset. Five candidate models were built, including one linear model and four generalized additive models (GAM). Candidate covariates (the seven cognitive tests available in both ADNI and GABS) were added to subsequent GAMs in order of theoretical similarity to the ADAS-Cog 11.

Rather than requiring a linear function to explain the relationship between the covariate(s) and outcome, GAMS allow a degree of nonlinearity in the dependence of the outcome on covariates.^{2,3} This is achieved by summing together smooth functions of covariates. Smooth functions are established for individual covariates by selecting a basis. A thin plate regression spline basis was used for all smooth functions in all candidate GAMS, and the basis dimension, which sets an upper limit to the number of degrees of freedom that the smooth function may take on, was set manually for each covariate based on visual

assessment of model plots. If a covariate was included in multiple candidate GAMs, the basis dimension was reassessed for each GAM as the amount of smoothing appropriate for a given covariate is influenced by other covariates in the model. In general, larger basis dimensions allow more degrees of freedom, which allows more nonlinearity in the smooth function. The amount of this “allowed” nonlinearity actually used (effective degrees of freedom (edf)) for a smooth function was selected through generalized cross validation as part of the standard model fit process in R. An edf of one indicates that a linear term was deemed acceptable by the GAM. R package `mgcv` was used to implement GAMs.⁴

The five candidate models were:

1. A linear model with the MMSE as the sole covariate. The MMSE is a global measure of cognition, shares many similar test items to the ADAS-Cog 11, and was selected as a starting point as it was expected to be the candidate covariate best able to independently predict ADAS-Cog 11 scores. The MMSE was included in all candidate models.
2. A GAM with the MMSE (basis=4, edf=2.7) as the sole covariate. This nonlinear model was superior to the previous linear model, so all subsequent candidate models were built as GAMs.
3. A GAM with the MMSE (basis=5, edf=3.7) and RAVLT (basis=10, edf=1.0) tests as covariates. The RAVLT is a test of episodic memory and resembles the ADAS-Cog 11 Word Recall task, which is one of the two items on the ADAS-Cog 11 where most errors are accumulated in pre-dementia populations. Due to this similarity, and the fact that the Word Recall task was identified as one of the most important ADAS-Cog 11 sub-items for assessing cognitive ability in pre-dementia populations, the RAVLT was the second covariate to be included.
4. A GAM with the MMSE (basis=6, edf=4.2), RAVLT (basis=10, edf=1.0), and CDR-SB (basis=7, edf=4.1) tests as covariates. The CDR-SB contains some similar assessments to the ADAS-Cog 11 (memory, orientation), but also includes report from a close relative, friend, or caregiver about functional activities (judgment, community affairs, home and hobbies, personal care items) that are not assessed directly by the ADAS-Cog 11.

5. A GAM with the MMSE (basis=6, edf=4.1), RAVLT (basis=10, edf=1.0), CDR-SB (basis=7, edf=4.1), TMT A (basis=5, edf=1.0), TMT B (basis=5, edf=1.0), Digit Span Forward (basis=5, edf=3.4), and Digit Span Backward (basis=5, edf=1.6) tests as covariates. The TMT parts A and B are both tests of executive function and processing speed, which are cognitive abilities not covered by the ADAS-Cog 11. The Digit Span Forwards and Backwards tests assess working memory and attention, which are also not directly assessed by the ADAS-Cog 11. All four of these tests covering additional cognitive domains were added at once to the final candidate GAM to assess how preliminary accuracy was changed by adding covariates that theoretically should not be very informative for estimating ADAS-Cog 11 scores.

Model 4 was selected as the best candidate model as it had better preliminary accuracy than Model 3, and only slightly worse preliminary accuracy than Model 5.

Candidate model diagnostics were assessed in the development subset to ensure model fit was okay. The corresponding plots can be found in Appendix B (Figures B.1 to B.5).

Accuracy estimation in testing data. Candidate Model 4 predicted 68.6% of ADAS-Cog 11 scores within 3 points of actual observed values (Figure 8, below), and 88.4% within 5 points. The Spearman rank correlation between predicted and observed ADAS-Cog scores was strong ($\rho=0.70$, $P<0.001$).

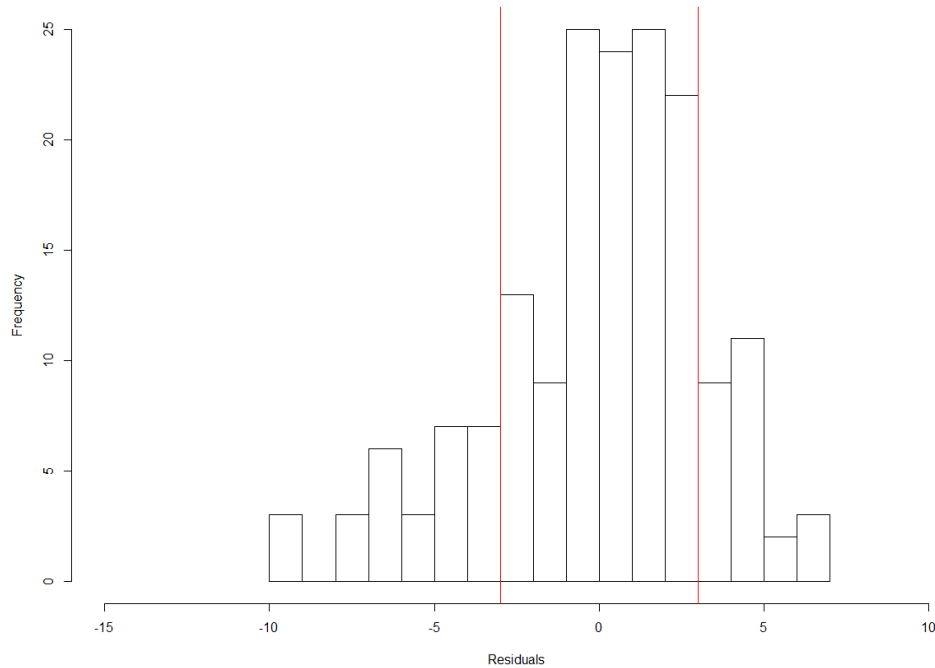


Figure 8 Residuals between observed and predicted ADAS-Cog scores in the ADNI testing data.

Comparison with previously published model. As a final check of Candidate Model 4 performance, the accuracy of a previously published univariate linear model for converting between MMSE and ADAS-Cog 11 scores was assessed.⁵ This model was developed in a sample of older adults with MCI and AD, which indicates higher levels of cognitive dysfunction than the ADNI or GABS sample, suggesting the model may not perform well enough for the purpose of approximating ADAS-Cog 11 scores in GABS. Indeed, this model predicted 53.2% of ADAS-Cog 11 scores within three points and 76.1% within five points of observed ADAS-Cog 11 scores on combined development and testing ADNI data.

Building the final ADAS-Cog-Proxy model. Candidate Model 4 was rebuilt on combined development and testing ADNI data, and is plotted in Figure 9.

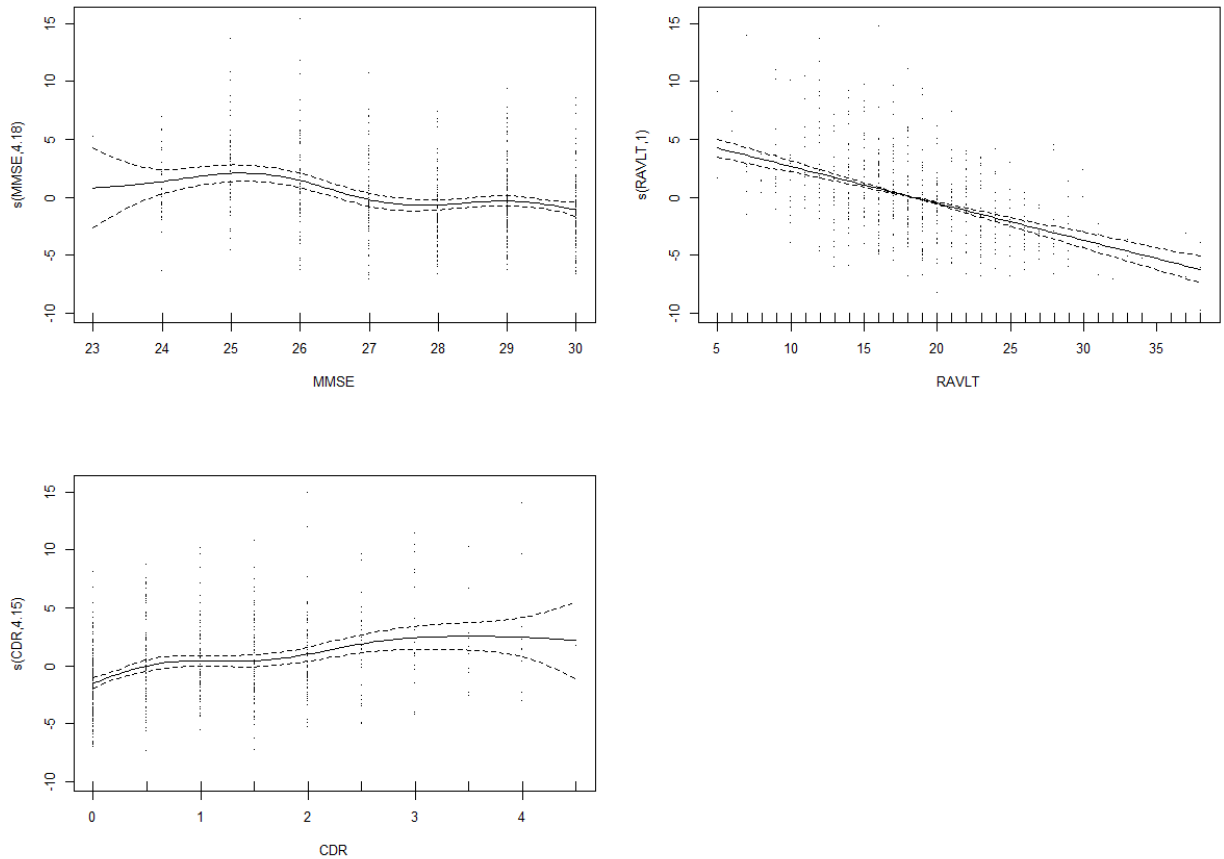


Figure 9 ADAS-Cog-Proxy model.

Shown are plots of the smooth terms (y-axis, number in brackets = effective degrees of freedom) against observed data points (x-axis) for each of the ADAS-Cog-Proxy generalized additive model covariates.

Assessing similarity of participants in ADNI and GABS. For optimal model performance, the participants in ADNI used to build the ADAS-Cog-Proxy model should be similar to the participants in GABS whom will be obtaining estimates of ADAS-Cog-Proxy scores. Given that both ADNI and GABS contain older adults along the pre-dementia disease continuum from NC to MCI, their cognitive abilities are expected to be similar. To ensure this was the case the range of ADAS-Cog-Proxy GAM covariates were compared between the ADNI data used to build the GAM, and observed GABS data, as presented in Table 13.

Table 13 Range of ADAS-Cog-Proxy Covariate Scores in ADNI and GABS

| Covariate Range (min, max) | MMSE 0, 30 | RAVLT 0, 45 | CDR-SB 0, 4.5 |
|---|---------------|----------------|------------------|
| ADNI data used to build ADAS-Cog-Proxy model | | | |
| Observed Range | 23, 30 | 5, 38 | 0, 4.5 |
| GABS Baseline | | | |
| Observed Range | 18, 30 | 8, 33 | 0, 4 |
| n below/above | 7/0 | 0 | 0 |
| Total n out of range | 7 | 0 | 0 |
| GABS 6-month follow-up | | | |
| Observed Range | 21, 30 | 5, 34 | 0, 4 |
| n below/above | 3/0 | 0 | 0/0 |
| Total n out of range | 3 | 0 | 0 |
| GABS 12-month follow-up | | | |
| Observed Range | 20, 30 | 6, 39 | 0, 2.5 |
| n below/above | 2 | 0/1 | 0/0 |
| Total n out of range | 2 | 1 | 0 |
| GABS 24-month follow-up | | | |
| Observed Range | 21, 30 | 9, 43 | 0.5, 5 |
| n below/above | 2/0 | 0/1 | 0/1 |
| Total n out of range | 2 | 1 | 1 |
| GABS 36-month follow-up | | | |
| Observed Range | 20, 30 | 6, 38 | 0.5, 3.5 |
| n below/above | 1/0 | 0 | 0/0 |
| Total n out of range | 1 | 0 | 0 |
| GABS 48-month follow-up | | | |
| Observed Range | 22, 30 | 5, 38 | 0.5, 4.0 |
| n below/above | 1/0 | 0 | 0/0 |
| Total n out of range | 1 | 0 | 0 |

Multiple Imputation by Chained Equations and ADAS-Cog-Proxy estimation in GABS. To allow all participants in GABS to obtain ADAS-Cog-Proxy scores, rather than omitting people who had at least one of three missing covariate values, Multiple Imputation by Chained Equations (MICE) was used (R package ‘mice’) to impute missing covariate values for all timepoints of interest.⁶⁻⁸

The number of missing ADAS-Cog-Proxy covariate values in GABS can be found in Table 14, below. CDR-SB missing values were coded as 999 if no collaborator, and NaN if unknown reason for missingness. This distinction in missingness was captured for Table 14, and then all missing values coded as NA so that MICE could be run.

Table 14 Missing ADAS-Cog-Proxy Covariates in the Gait and Brain Study

| Timepoint | n | MMSE | RAVLT | CDR-SB Total (no collaborator) | CDR-SB & RAVLT |
|-----------|-----|------|-------|--------------------------------------|----------------------|
| 0 | 109 | 0 | 29 | 68 (53) | 24 |
| 6m | 86 | 0 | 28 | 63 (50) | 25 |
| 12m | 73 | 0 | 16 | 57 (40) | 15 |
| 24m | 55 | 0 | 3 | 39 (33) | 3 |
| 36m | 35 | 0 | 0 | 21 (21) | 0 |
| 48m | 24 | 0 | 0 | 20 (19) | 0 |

Legend: MMSE=Mini-Mental State Examination, RAVLT=Rey Auditory Verbal Learning Test (3 trials), CDR-SB=Clinical Dementia Rating Scale-Sum of Boxes, m=months.

The imputation method of predictive mean matching (default) was used to impute missing values for both CDR-SB and RAVLT. Predictive mean matching is a semi-parametric imputation method which generates a prediction for missing values using other variables in the predictor matrix, and then selects an observed value from the predictor matrix that is similar to the predicted value. The default visit sequence of imputing variables in order from left to right was used, and five multiply imputed datasets were created.

MICE was performed on extracted predictor matrices whereby only the final pooled ADAS-Cog-Proxy scores were merged back into the original GABS dataset, rather than performing MICE on the entire GABS database, due to multicollinearity and computational restrictions.^{6,7} The creation of predictor matrices for each time point also allowed the exclusion of observations that were missing simply because the corresponding participants did not have the follow-up visit. Furthermore, it has been suggested that there is no advantage in terms of accuracy for imputations when using more than 15-25 predictor variables.⁶ In accordance with published guidelines, predictor matrices included all GAM covariates, predictors of the outcome ADAS-Cog scores, variables that include a lot of variance as roughly identified by correlation with the target variables to be imputed (Table 15), and no variables that had a lot of missing values within the subgroup of people with missing RAVLT and CDR-SB scores.⁶⁻⁸ It has also been suggested to include variables related to non-response. The main reason CDR-SB

scores are missing is if no collaborator was present to report on behalf of the patient; however, there was not a variable in the dataset expected to provide indication of this.

Table 15 Correlation Coefficients for Potential Predictor Matrix Variables

| Candidate Predictor Variable | Pearson Correlation Coefficient with CDR | Pearson Correlation Coefficient with RAVLT | Include variable in predictor matrix? |
|---|--|--|---------------------------------------|
| Activities of Daily Living (ADL) | NA | -0.12 | No |
| Age (years) | 0.07 | -0.23 | No |
| Balance-Rigid Surface Eyes Open | 0.28 | -0.09 | Yes |
| Balance-Rigid Surface Eyes Shut | 0.11 | -0.17 | No |
| Balance-Disturbed Surface Eyes Open | 0.03 | -0.11 | No |
| Balance-Disturbed Surface Eyes Shut | 0.06 | -0.06 | No |
| Basic ADL | 0.05 | 0.02 | No |
| Boston Naming Test | -0.40 | 0.34 | Yes |
| Clinical Dementia Rating Scale | 1.00 | -0.46 | Yes |
| Diagnostic Category | NA | NA | Yes |
| Digit Forward Span | 0.24 | 0.20 | Yes |
| Digit Backward Span | -0.21 | 0.20 | Yes |
| Education (years) | -0.11 | 0.03 | No |
| Frontal Assessment Battery | -0.55 | 0.28 | Yes |
| Gait Velocity | -0.38 | 0.17 | Yes |
| Gait Velocity while Counting | -0.29 | 0.22 | Yes |
| Geriatric Depression Scale | 0.03 | -0.05 | No |
| Instrumental ADL | -0.37 | 0.07 | Yes |
| Letter Number | -0.32 | 0.29 | Yes |
| Mini-Mental State Examination | -0.56 | 0.28 | Yes |
| Montreal Cognitive Assessment (MoCA) | -0.67 | 0.38 | Yes |
| MoCA Attention Index Score | -0.26 | 0.25 | Yes |
| MoCA Executive Index Score | -0.46 | 0.20 | Yes |
| MoCA Language Index Score | -0.26 | 0.29 | Yes |
| MoCA Memory Index Score | -0.48 | 0.41 | Yes |
| MoCA Orientation Index Score | -0.64 | 0.27 | Yes |
| MoCA Visuospatial Index Score | -0.27 | 0.02 | Yes |
| Number of falls in the past 12 months | -0.26 | 0.05 | Yes |
| Physical Activity Scale for the Elderly | -0.09 | -0.09 | No |
| Rey Auditory Verbal Learning Test | -0.46 | 1.00 | Yes |
| Trail Making Task A | 0.28 | -0.44 | Yes |
| Trail Making Task B | 0.49 | -0.25 | Yes |

The following variables were included in predictor matrices for each time point: Baseline Diagnosis, MMSE, MoCA, MoCAMIS, MoCAEIS, MoCAVIS, MoCALIS, MoCAAIS, MoCAOIS, CDR, Trail A, Trail B, Digit Forward, Digit Backward, Letter Number,

RAVLT, BNT, FAB, number of falls in past 6 months, IADL, RSEO (balance), Gait Velocity, and Gait Velocity while counting backwards by ones, from the time point of interest, as well as CDR and RAVLT scores from the previous visit (T6 to T48 visit imputations) or a future visit (baseline visit imputations). Participant ID was included in the predictor matrix to allow re-merging of data, but was omitted as a predictor variable.

After model specification and predictor matrix creation, MICE was performed in 3 main stages,^{6,8} and repeated for each timepoint.

1) The imputation stage created five multiply imputed datasets. Although only CDR-SB and RAVLT imputations were required, to remove a variable from being imputed it must also be removed as a predictor variable, so imputations were allowed for predictor variables that had missing values themselves. Imputations for RAVLT and CDR-SB were inspected visually to ascertain the plausibility of imputed values. Convergence of the MICE algorithm was also assessed by plotting imputations streams for the mean and standard deviation (y-axes) of the five imputations against the iteration number (x-axes), as shown in Figures B.6 to B.11 in Appendix B. Imputation streams that are intermingled without definite trends may be considered as support for convergence.⁶

2) The analysis stage included applying the ADAS-Cog-Proxy GAM to each of the five complete datasets, which resulted in five estimated ADAS-Cog-Proxy scores for each participant in GABS.

3) The pooling stage involved taking the mean of the five estimated ADAS-Cog-Proxy scores for each participant. These final averaged scores were labelled “T#_ADASproxy” and merged back into the GABS database using Participant ID.

A summary of ADAS-Cog-Proxy descriptive statistics at each time point were assessed to ensure merging was performed correctly and to further assess plausibility of the estimated scores, and are presented in Table 16.

Table 16 ADAS-Cog-Proxy Scores in the Gait and Brain Study

| Timepoint | Minimum | Quartile 1 | Median | Mean | Quartile 3 | Maximum |
|-----------|---------|------------|--------|-------|------------|---------|
| Baseline | 2.77 | 7.89 | 9.34 | 9.74 | 11.36 | 18.41 |
| 6 Months | 3.80 | 8.24 | 9.61 | 9.75 | 11.36 | 17.24 |
| 12 Months | 2.60 | 7.34 | 8.78 | 9.24 | 11.04 | 17.89 |
| 24 Months | 0.29 | 6.85 | 8.43 | 8.61 | 10.62 | 14.03 |
| 36 Months | 2.75 | 7.93 | 9.62 | 9.77 | 10.88 | 17.30 |
| 48 Months | 5.26 | 9.05 | 10.64 | 10.95 | 12.36 | 17.08 |

6.1.2 Step 2: Select Additional Measures for the Novel Outcome Measure

The overall goal for PI component selection was to select lowly correlated variables, with at least one variable from each of the following three categories thought to be important components of pre-dementia and dementia syndromes: cognition, motor function, and motor-cognitive performance. For simplicity candidate variables were separated into these three categories, but in reality these categories are not mutually exclusive. Including up to six component variables with low pairwise correlations in a PI has been shown to be advantageous in terms of content validation (covering multiple important domains) and in terms of reducing the variability of the final PI score.⁹⁻¹¹ Although statistically there is little or no advantage to including more than six component variables or component variables with pairwise correlations higher than about $|\rho|=0.4$ in a PI, doing so does not affect the validation of the PI nor preclude potential non-statistical advantages that may be gained by including certain measures.⁹

To select candidate variables to include in the PI the following steps were followed, using baseline GABS data. Correlation coefficients were calculated using all complete pairwise correlations.

1. Treat the ADAS-Cog-Proxy as the “base” of the PI to cover the cognitive domain.
2. All single-task quantitative gait parameters gathered by the GAITRite™ electronic walkway system were considered for the motor function category of potential PI components. Pairwise correlations were calculated between all quantitative gait parameters and the ADAS-Cog-Proxy, Table 17:

Table 17 Correlation Coefficients for Quantitative Gait Parameters and ADAS-Cog-Proxy

| Single-Task Gait Parameter | Correlation with ADAS-Cog-Proxy |
|---------------------------------|---------------------------------|
| Gait Velocity | -0.32 |
| Stride Time | 0.26 |
| Stride Time Variability | 0.09 |
| Step Time | 0.25 |
| Stride Length | -0.28 |
| Step Length | -0.28 |
| Double Support Time | 0.03 |
| Swing Time | 0.04 |
| Stride Width | -0.06 |
| Stride Velocity | -0.33 |
| Cadence | -0.26 |
| Step Time Variability | 0.08 |
| Stride Length Variability | 0.11 |
| Step Length Variability | 0.13 |
| Double Support Time Variability | 0.12 |
| Swing Time Variability | 0.16 |
| Stride Width Variability | -0.001 |
| Stride Velocity Variability | 0.07 |

3. Nine Dual-Task Cost (DTC) assessments were selected to be included in the functionality category of candidate PI components based on the presence of literature supporting their importance for dementia or pre-dementia syndromes. Details on the DTC paradigm can be found in Chapter 5. Pairwise correlations were calculated between candidate DTC variables and the ADAS-Cog-Proxy, Table 18:

Table 18 Correlation Coefficients for Dual-Task Cost Assessments and ADAS-Cog-Proxy

| Dual-Task Cost (Secondary Task, Gait Parameter) | Correlation with ADAS-Cog-Proxy |
|---|---------------------------------|
| Counting Backwards by Ones, Gait Velocity | 0.28 |
| Counting Backwards by Ones, Stride Time | -0.28 |
| Counting Backwards by Ones, Stride Time Variability | -0.30 |
| Counting by Serial Sevens, Gait Velocity | 0.28 |
| Counting by Serial Sevens, Stride Time | -0.22 |
| Counting by Serial Sevens, Stride Time Variability | -0.26 |
| Naming Animals, Gait Velocity | 0.30 |
| Naming Animals, Stride Time | -0.27 |
| Naming Animals, Stride Time Variability | -0.14 |

4. Assess pairwise correlations between single task gait and DTC variables that, in Steps 2 and 3, had $|\rho| < 0.2$ with the ADAS-Cog-Proxy, or had $|\rho| = 0.2$ to 0.4 with the ADAS-Cog-Proxy and evidence in published literature demonstrating importance in dementia or predementia syndromes, or demonstrating significant associations with cognition or functionality abilities thought to be important for older adults with dementia or pre-dementia syndromes (Table 19). It was also ensured that the direction of correlation coefficients was congruent with the ADAS-Cog-Proxy scoring of higher indicating worse dysfunction. If a variable was scored as higher numbers indicating worse performance, positive correlation coefficients were favoured. If a variable was scored as higher scores indicate less dysfunction, negative correlation coefficients were favoured.

Table 19 Correlation Coefficients for Single Task and Dual-Task Cost Gait Variables that were Retained after Steps One to Three

| | GV | ST | STV | SWT | SW | DSTV | SWV | SVV | C_GV | 7_GV | A_GV | A_STV |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| GV | 1.00 | -0.63 | -0.28 | -0.09 | -0.09 | -0.01 | 0.06 | -0.45 | -0.27 | -0.28 | -0.25 | 0.11 |
| ST | -0.63 | 1.00 | 0.17 | 0.67 | 0.07 | -0.05 | -0.11 | 0.24 | 0.13 | 0.28 | 0.22 | -0.19 |
| STV | -0.28 | 0.16 | 1.00 | 0.04 | -0.09 | 0.15 | 0.20 | 0.55 | 0.17 | 0.07 | 0.02 | 0.25 |
| SWT | -0.09 | 0.67 | 0.04 | 1.00 | -0.19 | -0.18 | -0.02 | 0.06 | 0.02 | 0.21 | 0.10 | -0.18 |
| SW | -0.09 | 0.07 | -0.09 | -0.19 | 1.00 | 0.11 | -0.43 | -0.02 | -0.09 | -0.01 | 0.03 | -0.07 |
| DSTV | -0.01 | -0.05 | 0.15 | -0.18 | 0.11 | 1.00 | -0.05 | 0.36 | 0.17 | 0.10 | 0.07 | 0.02 |
| SWV | 0.06 | -0.11 | 0.20 | -0.02 | -0.43 | -0.05 | 1.00 | 0.09 | 0.08 | -0.12 | -0.03 | 0.12 |
| SVV | -0.45 | 0.24 | 0.55 | 0.06 | -0.2 | 0.36 | 0.09 | 1.00 | 0.21 | 0.13 | 0.15 | 0.08 |
| C_GV | -0.27 | 0.13 | 0.17 | 0.02 | -0.09 | 0.17 | 0.09 | 0.21 | 1.00 | 0.64 | 0.71 | -0.28 |
| 7_GV | -0.28 | 0.28 | 0.07 | 0.21 | -0.01 | 0.10 | -0.12 | 0.13 | 0.64 | 1.00 | 0.82 | -0.45 |
| A_GV | -0.25 | 0.22 | 0.02 | 0.10 | 0.03 | 0.07 | -0.03 | 0.15 | 0.71 | 0.82 | 1.00 | -0.56 |
| A_STV | 0.11 | -0.19 | 0.25 | -0.18 | -0.07 | 0.02 | 0.12 | 0.08 | -0.28 | -0.45 | -0.56 | 1.00 |

Legend: GV=Gait Velocity, ST=Stride Time, STV=Stride Time Variability, SWT=Swing Time, SW=Stride Width, Double Support Time Variability, Stride Width Variability, SVV=Stride Velocity Variability, C_GV=Dual-Task Cost (DTC) for Gait Velocity with Counting backwards by ones, 7_GV=DTC for Gait Velocity while counting backwards by Serial Sevens, A_GV=DTC for Gait Velocity while Naming Animals, A_STV=DTC for Stride Time Variability while Naming Animals.

5. When numerical and theoretical considerations were similar, as described in Step 4, box plots of the individual gait (Figure 10 to 12) or DTC (Figure 13 to 16) variables were created to visually assess distribution between diagnostic

categories of NC, SCI, and MCI. Variables showing greater variability and a stepwise progression from NC to SCI to MCI were favoured.

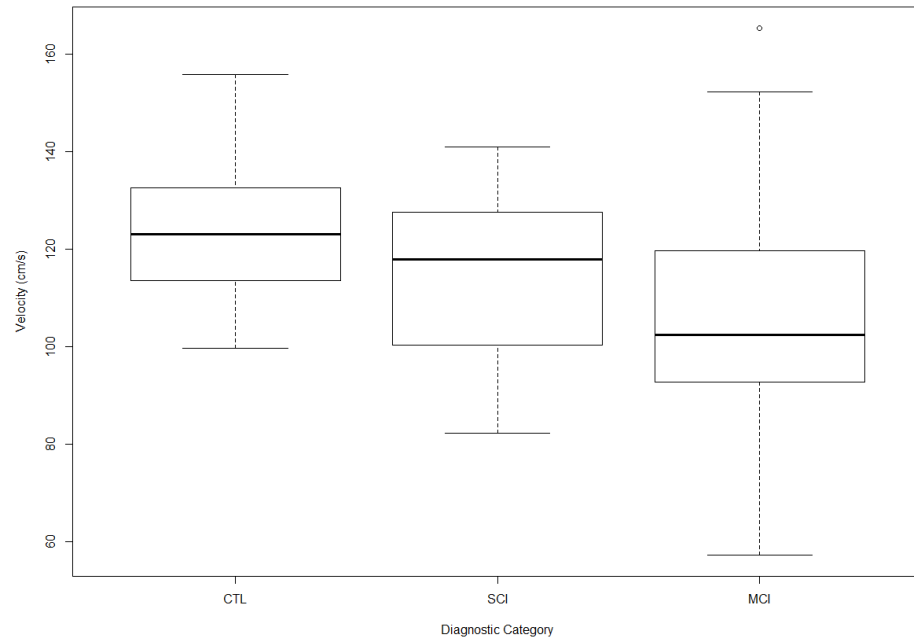


Figure 10 Box plot of gait velocity by baseline diagnostic category.

Legend: CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment.

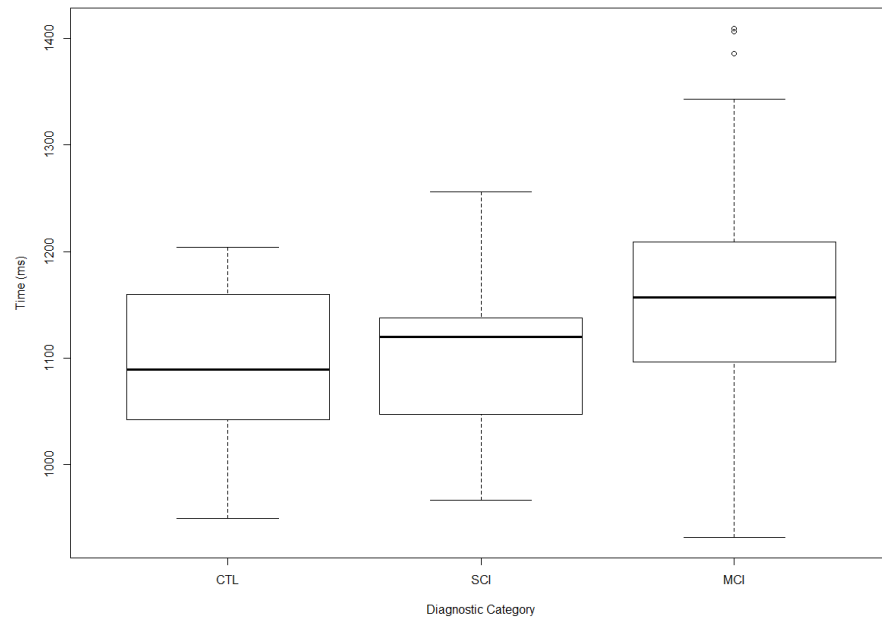


Figure 11 Box plot of gait stride time by baseline diagnostic category.

Legend: CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment.

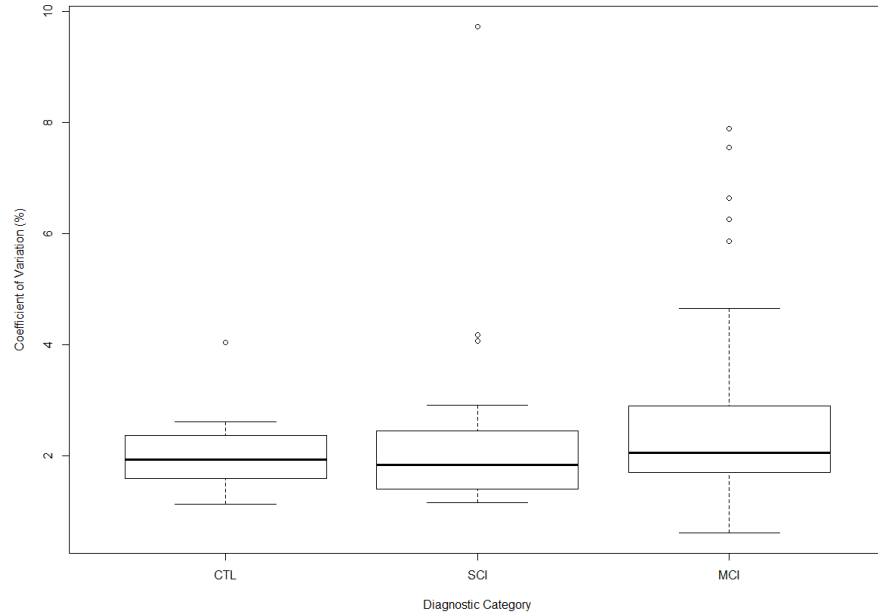


Figure 12 Box plot of gait stride time variability by baseline diagnostic category.
Legend: CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment.

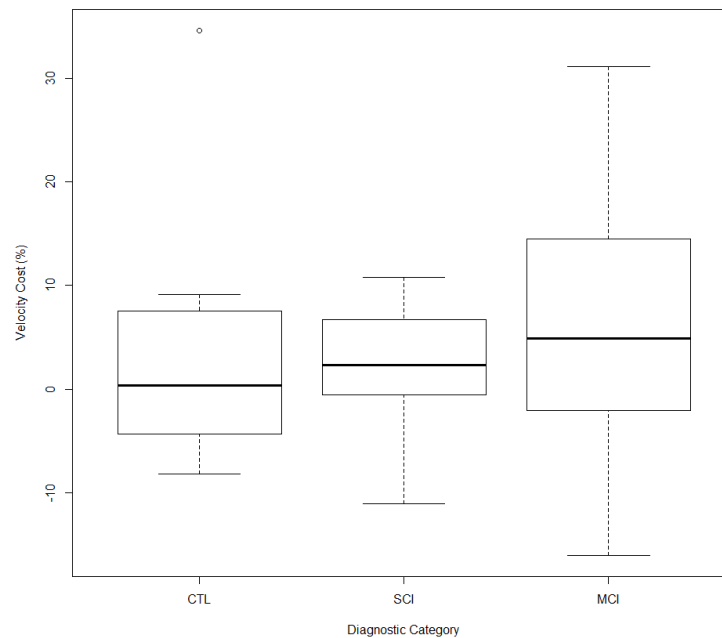


Figure 13 Box plot of dual-task gait velocity cost when counting backwards by ones against baseline diagnostic category.
Legend: CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment.

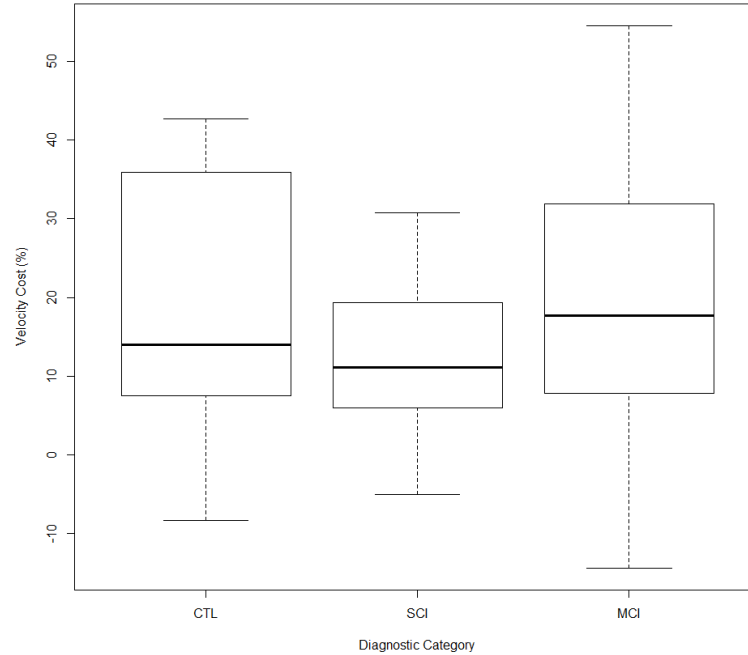


Figure 14 Box plot of dual-task gait velocity cost when counting backwards by serial sevens against baseline diagnostic category.

Legend: CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment.

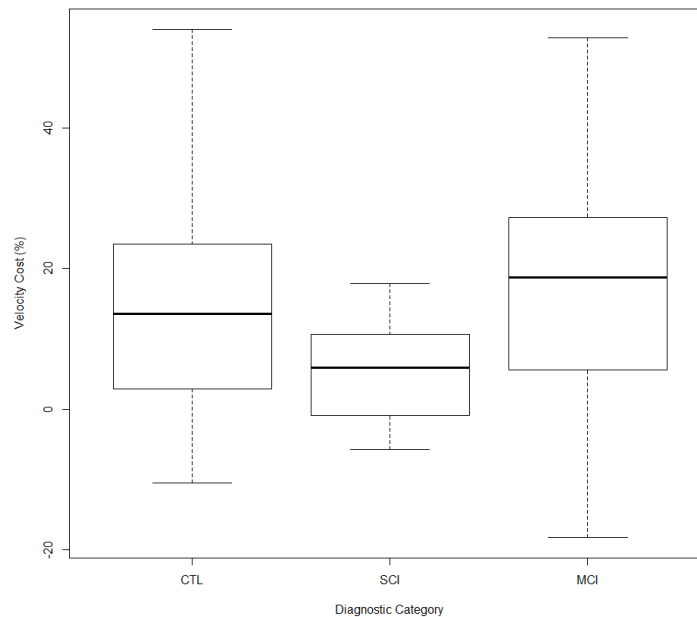


Figure 15 Box plot dual-task gait velocity cost when naming animals against baseline diagnostic category.

Legend: CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment.

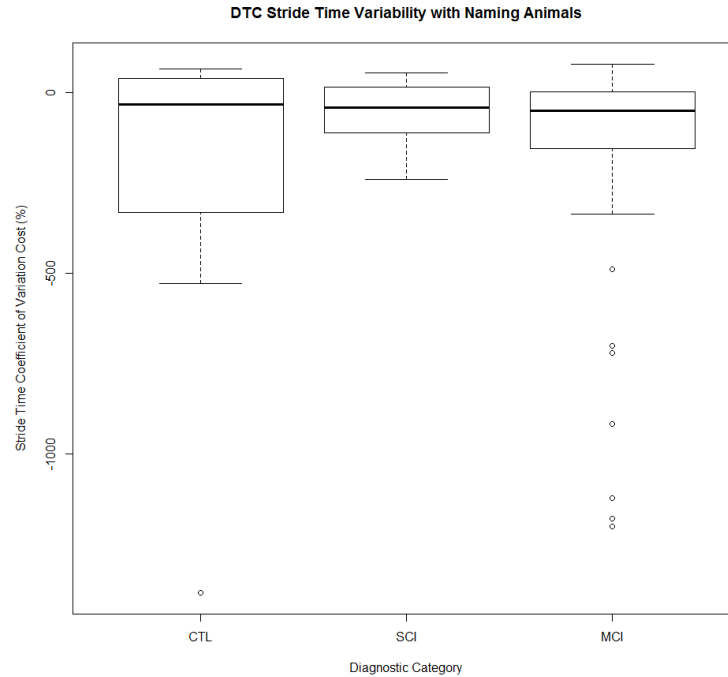


Figure 16 Box plot of dual-task stride time variability cost when naming animals against baseline diagnostic category.

Legend: CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment.

6. When numerical, theoretical, and distributional assessments were similar, practical considerations in terms of ease of assessment were taken into account both for individual variables and for the PI as a whole. Gait velocity is the only quantitative gait parameter that can be measured easily without the use of an electronic gait mat.
7. Scatterplots of the most promising candidate variables and the ADAS-Cog-Proxy were assessed to ensure low pairwise correlation coefficients were not in spite of a strong non-linear relationship, Figures 17 to 19:

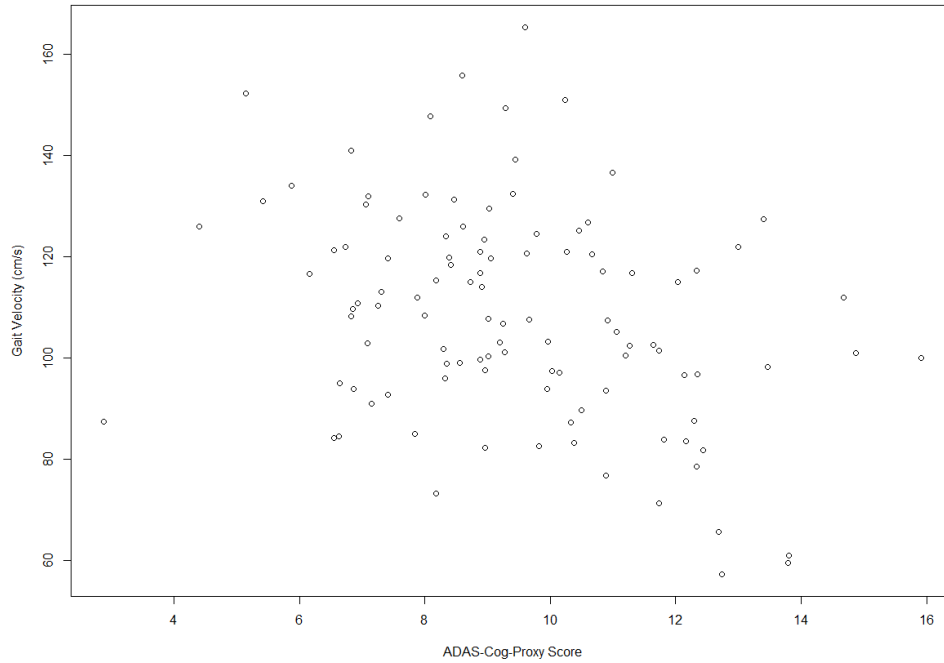


Figure 17 Scatterplot of gait velocity against ADAS-Cog-Proxy scores.

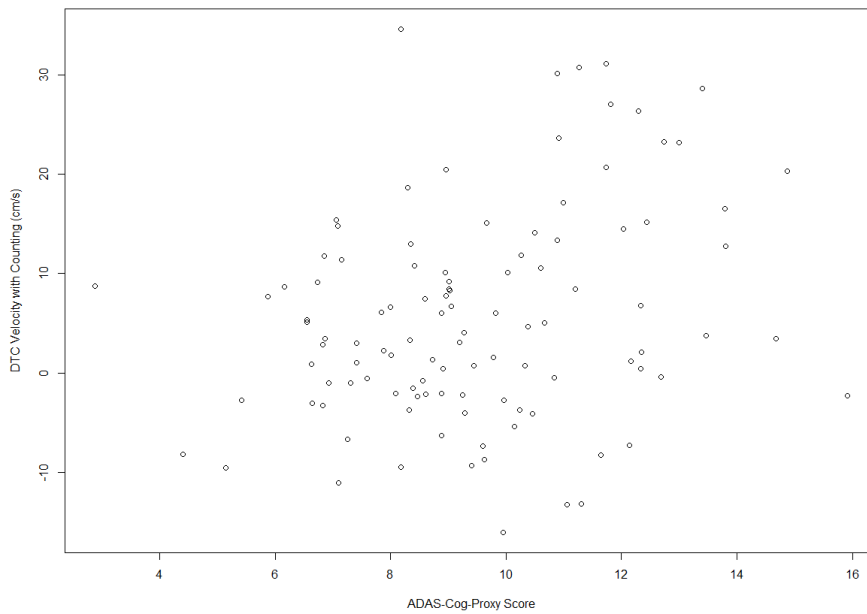


Figure 18 Scatterplot of dual-task gait velocity cost with secondary task of counting backwards by ones against ADAS-Cog-Proxy scores.

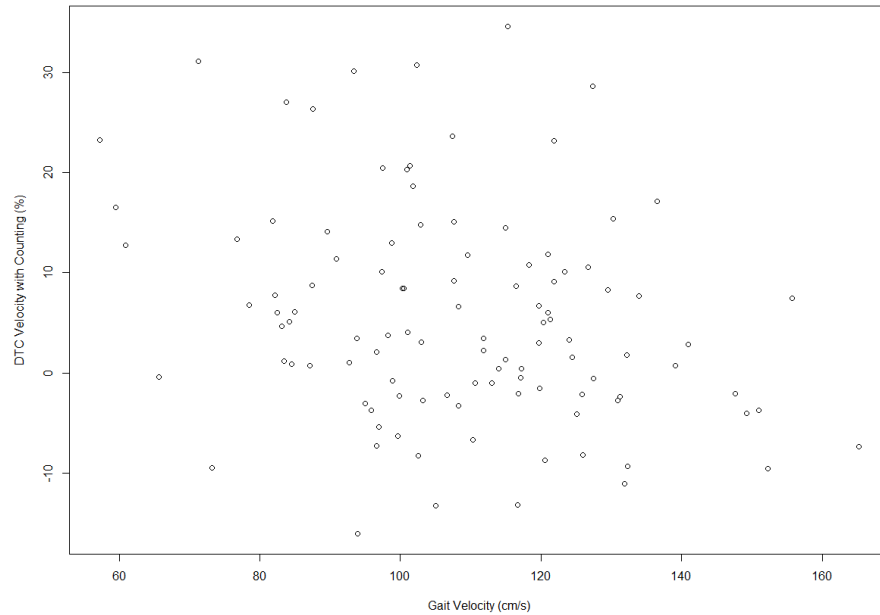


Figure 19 Scatterplot of dual-task gait velocity cost with secondary task of counting backwards by ones against gait velocity.

The final three variables selected for inclusion in the PI were ADAS-Cog-Proxy, gait velocity, and DTC for gait velocity with the secondary task of counting backwards from 100 by ones. Pairwise correlation coefficients are presented below, in Table 20.

Table 20 Pairwise Correlation Coefficients between Final Pooled Index Components

| | ADAS-Cog-Proxy | Gait Velocity | Dual-Task Cost |
|----------------|----------------|---------------|----------------|
| ADAS-Cog-Proxy | 1.00 | -0.32 | 0.28 |
| Gait Velocity | -0.32 | 1.00 | -0.27 |
| Dual-Task Cost | 0.28 | -0.27 | 1.00 |

6.1.3 Step 3: Combine Measures Using a Pooled Index Approach

Baseline PI scores (for Objective 2) were calculated according to the following steps⁹⁻¹¹:

- 1) Multiply gait velocity by -1 so that all variables are coded as higher scores indicating worse performance (slower=worse).

- 2) Calculate the baseline mean and Standard Deviation (SD) separately for gait velocity, DTC, and ADAS-Cog-Proxy measures.
- 3) Calculate standardized scores ($Z = (X_i - \bar{X})/SD$) for gait velocity, DTC, and ADAS-Cog-Proxy measures.
- 4) Sum together the three standardized scores and divide by three (take average).

6.2 Objective 2

A One-Way Analysis of Variance (ANOVA) was the initially planned statistical test to assess responsiveness of the ADAS-Cog-Proxy and of the PI to group-level between-person differences in stage of disease progression at one point in time. Box plots and QQ plots, included in Appendix B (Figures B.12 to B.15), were used to assess the suitability of parametric tests. Assumptions of normality and homoscedasticity were not supported, and there were small sample sizes for NC and SCI categories, so the non-parametric Kruskal Wallis test was used instead of an ANOVA to assess whether the three diagnostic categories arose from the same distribution. Non-parametric Mann-Whitney U tests were then used to assess all pairwise comparisons between diagnostic categories. Results were presented in Chapter 5.

Including Activities of Daily Living. An additional analysis not included in Chapter 5 includes a Lawton-Brody Instrumental Activities of Daily Living (IADL) test. Although the GABS sample contains strictly participants with pre-dementia stages of disease progression where by diagnostic definition ADLs must be intact,¹² ADLs are a common assessment of functionality and have previously been shown to improve responsiveness of the ADAS-Cog 11.¹³ To assess whether it would be beneficial to include an ADL measure on top of the motor, DTC, and cognitive measures included in our PI, a four component PI was created that included the cognitive, motor function, and DTC variables as well as the Lawton Brody IADL assessment (score reversed). Visual evaluation was used to compare baseline box plots for the four component PI (Figure 20, below) and the three component PI (Figure 21, below). The addition of the IADL assessment did not

appear to provide an advantage for baseline discrimination, so the simpler PI version was maintained.

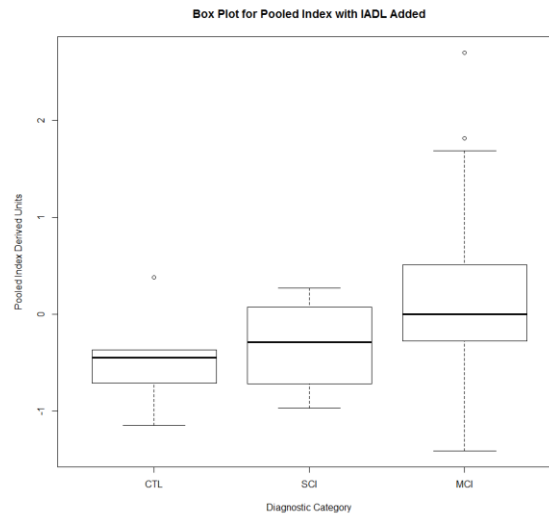


Figure 20 Box plot of a pooled index that includes an assessment of instrumental activities of daily living against baseline diagnostic category.

Pooled index components included: ADAS-Cog-Proxy, gait velocity, DTC gait velocity with secondary task of counting, and instrumental activities of daily living assessments.

Legend: CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment.

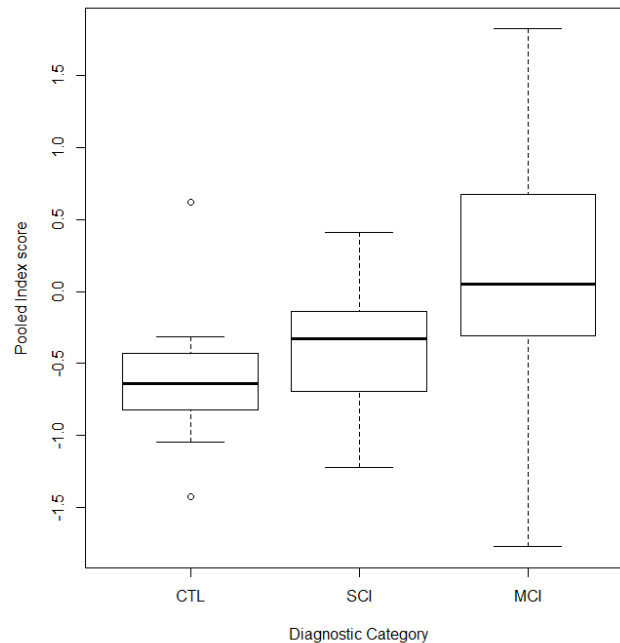


Figure 21 Box plot of pooled index against baseline diagnostic category.

Pooled index components included: ADAS-Cog-Proxy, gait velocity, and DTC gait velocity with secondary task of counting. **Legend:** CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment.

6.3 Objective 3

Standardized Response Mean (SRM) calculations were performed for the ADAS-Cog-Proxy, the complete PI, and the standardized ADAS-Cog-Proxy combined with each individual component of the PI (also using a PI approach) for 6, 12, 24, 36, and 48 month follow-up periods. Larger SRM values indicate better responsiveness to measured decline over time, and can be calculated with the following formula¹⁴:

$$\text{SRM} = \text{mean}(\text{difference score}) / \text{SD}(\text{difference score})$$

PI difference scores (both complete and subcomponent versions) were calculated according to the following steps:

1. Create a variable for DTC at each timepoint.
2. Create a variable for reversed gait velocity (multiply by -1) at each time point. Summary statistics were calculated to check that the reversal was performed correctly.
3. Create subsets of data that correspond to the participants present at each follow-up time point. This will allow standardization with respect to the baseline distribution of participants who have been enrolled in the study long enough to reach the desired follow-up timepoint without dropping out or being excluded.

Steps 4 to 10 were completed for each follow up subset of data.

4. For each variable, calculate the baseline mean for the subset of participants present at the follow-up visit.
5. For each variable, calculate the baseline SD for the subset of participants present at the follow-up visit.
6. For each variable, calculate a standardized baseline score as:

$$Z = (X_{i,baseline} - \bar{X}_{baseline}) / SD_{baseline}$$

7. Calculate baseline PI scores by taking the average of the Z-scores calculated in Step 6.
8. For each variable, calculate a standardized follow-up score as:

$$Z = (X_{i, follow-up} - \bar{X}_{baseline}) / SD_{baseline}$$

9. Calculate follow-up PI scores by taking the average of the Z-scores calculated in Step 8.
10. To get difference scores, subtract the baseline PI score (Step 7) from the follow-up PI score (Step 9).

All SRM results can be found in Chapter 5, Table 11.

6.4 References

1. Tuszynski J, Khachatryan H. Package “ caTools .” 2015. <http://cran.r-project.org/package=caTools>.
2. Wood SN. *Generalized Additive Models: An Introduction with R*. Boca Raton: Taylor & Francis Group; 2006.
3. Hastie T, Tibshirani R. Generalized additive models. *Stat Sci*. 1986;1(3):297-318. doi:10.1016/j.csda.2010.05.004.
4. Wood S. Package ‘mgcv’. 2017. <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>
5. Solomon TM, DeBros GB, Budson AE, Mirkovic N, Murphy CA, Solomon PR. Correlational analysis of 5 commonly used measures of cognitive functioning and mental status: an update. *Am J Alzheimers Dis Other Demen*. 2014;29(8):718-722. doi:10.1177/1533317514534761.
6. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3). doi:10.18637/jss.v045.i03.
7. van Buuren S, Oudshoorn K. *Flexible Multivariate Imputation by MICE*. Netherlands Organization for Applied Science Research;1999.
8. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30(4):377-399. doi:10.1002/sim.4067.
9. Smythe H, Helewa A, Goldsmith C. “Independent assessor” and “pooled index” as techniques for measuring treatment effects in rheumatoid arthritis. *J Rheumatol*. 1977;4(2):144-152.
10. Goldsmith C, Smythe H, Helewa A. Interpretation and power of a pooled index. *J Rheumatol*. 1993;20(3):575-578.
11. Speechley M, Forchuk C, Hoch J, Jensen E, Wagg J. Deriving a mental health outcome measure using the pooled index: an application to psychiatric consumer-survivors in different housing types. *Heal Serv Outcomes Res Methodol*. 2009;9(2):133-143. doi:10.1007/s10742-009-0044-4.
12. Petersen RC. Mild Cognitive Impairment. *N Engl J Med*. 2011;364:2227-2234. doi:10.1017/S1355617714000137.
13. Wessels A, Siemers E, Andersen S, et al. A combined measure of cognition and function for clinical trials: the integrated Alzheimer’s Disease Rating Scale (iADRS). *J Prev Alzheimers Dis*. 2015;2(4):227-241. doi:10.1007/978-1-4614-5915-6.
14. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness. *J Clin Epidemiol*. 2000;53(5):459-468. doi:10.1016/S0895-4356(99)00206-1.

Chapter 7

7 Discussion

The purpose of Chapter 6 is to provide an overview of the findings of this thesis, put these findings into the context of peer-reviewed literature, discuss limitations, suggest directions for future research, and highlight clinical implications.

7.1 Summary of Findings

The purpose of this thesis was to assess use of the present gold standard for assessing efficacy of antidementia treatments, namely the Alzheimer's Disease (AD) Assessment Scale–Cognitive Subscale (ADAS-Cog 11), in pre-dementia populations, and to explore whether adding motor function assessments to the ADAS-Cog 11 improves responsiveness in a pre-dementia sample.

The literature review (Chapter 3) suggested that the ADAS-Cog 11 is not optimally responsive for use in pre-dementia populations. Furthermore, its content validity is reduced through research advancements that propose constructs not assessed by the ADAS-Cog 11, such as executive function (EF) and motor function, that could be understood as important aspects of disease severity. Several modification approaches have improved the responsiveness of the ADAS-Cog 11 through altered scoring methodology or content, but no modifications incorporating motor function were found. This may be, at least in part, due to the apparent lack of a database that contains both the ADAS-Cog 11 and high-quality motor function assessments. An ADAS-Cog-Proxy measure was developed to address this challenge. The corresponding methods, including concurrent criterion validation of the ADAS-Cog-Proxy model with the ADAS-Cog 11 in the Alzheimer's Disease Neuroimaging Initiative database, may be applied in similar situations where all desired variables are not present in a single database, but one of the key variables is present in a second database with comparable subjects. Both databases must contain overlapping variables that are similar to the key variable. After approximation of ADAS-Cog 11 scores in the Gait and Brain Study (GABS), a pooled

index (PI) approach was used to combine quantitative motor assessments with ADAS-Cog-Proxy scores. The resulting Gait and Brain Pooled Index (GAB-PI) consists of gait velocity, dual-task gait velocity cost (DTC), and the ADAS-Cog-Proxy.

Overall the GAB-PI showed similar levels of responsiveness as the ADAS-Cog-Proxy to baseline discrimination of pre-dementia syndromes, demonstrating that combining the ADAS-Cog 11 with motor function assessments did not obscure its ability to detect differences between diagnostic categories that were defined based on a primarily cognitive conceptualization of the disease. Specifically, both the ADAS-Cog-Proxy and GAB-PI demonstrated the ability to discriminate between normal cognition (NC) and Mild Cognitive Impairment (MCI), and between MCI and Subjective Cognitive Impairment (SCI), but not between NC and SCI groups. Other studies have found similar results for the ADAS-Cog 11, and for ADAS-Cog 11 modifications.¹⁻⁴

Although diagnostic categories can serve many useful purposes, they do not necessarily represent all meaningful changes that may occur during progression, or regression, within diagnostic categories. Analyses of responsiveness to disease progression whereby no diagnostic classification was used, except to exclude older adults who progressed to dementia, found that in general the GAB-PI had comparable or slightly better responsiveness than the ADAS-Cog-Proxy. Specifically, for six and forty-eight month follow-up periods the GAB-PI detected more decline than the ADAS-Cog-Proxy, while for thirty-six months of follow-up the ADAS-Cog-Proxy detected greater decline. For twelve and twenty-four month periods the ADAS-Cog-Proxy found overall improvement while the GAB-PI found almost no change. Although not common, other studies that included older adults with NC or MCI have found improvement over one to two years on the ADAS-Cog 11 as well as on ADAS-Cog modifications (ADAS-Cog 12, ADAS-Cog 13, ADAS-Cog 3, ADAS-Cog 5, ADAS-Rasch, ADNI-Mem, ADAS-Tree).^{5,6} Treating the ADAS-Cog-Proxy as the gold standard, and the cognitive modifications as potential improvements thereof, this indicates that the GAB-PI may not be picking up on changes in the correct direction. However, in one of the studies where the ADAS-Cog 11 and modified versions found improvement over two-years for older adults with NC, the Clinical Dementia Rating Scale-Sum of Boxes (CDR-SB), which is a well-respected

assessment of overall disease severity,⁷⁻⁹ found the largest magnitude of change over time, and it was in the direction of worsening.⁵ Furthermore, knowing that the progression from NC to MCI to dementia is not linear,¹⁰ but that motor function decline may be detectable in advance of further cognitive decline,^{11,12} and that motor function and cognitive ability are both important for functionality,¹³⁻¹⁶ it is possible that the measures detecting decline over time are presenting a more accurate detection of changes in overall disease severity. Further validation analyses against clinical indicators of overall disease severity may help to clarify this.

Interestingly, the PI combining only gait velocity with the ADAS-Cog-Proxy increased responsiveness to measured decline compared to the ADAS-Cog-Proxy alone for all follow-up periods, and demonstrated larger responsiveness than the full GAB-PI for the two longest follow-up periods (36 and 48 months). The PI combining only DTC with the ADAS-Cog-Proxy was never more responsive than the full GAB-PI, and for at least two follow-up periods was less responsive than the ADAS-Cog-Proxy alone.

These findings were surprising as the DTC paradigm was designed to tap into the limited capacity of the brain such that when one is asked to perform an attention demanding task while walking, some of the brain capacity that would otherwise be devoted to walking (single-task gait can use all available capacity) is reallocated to the secondary task, and gait performance worsens.^{12,17-20} As neurodegeneration progresses en route to dementia, there is less and less capacity to devote to the two tasks, and the DTC should increase.^{14,21,22} Previous research on gait assessments alone has found significant differences between single- and dual-task gait parameters within MCI and NC populations,^{18,23-26} whereby the disruption is often greater for MCI than NC.^{17,24,26,27} The ability of DTC to discriminate between pre-dementia syndromes is better than single-task gait,^{27,28} and DTC in MCI has been associated with risk of progression to dementia.²⁹ Hence, more consistent and advantageous results from the addition of a DTC assessment to the ADAS-Cog-Proxy for responsiveness to disease progression was expected.

The cueing effect refers to an improvement in gait under dual-task compared to single-task conditions, and is thought to occur because the secondary task can serve as a pacer or rhythmic auditory stimulation to aid gait.³⁰ The cueing effect, may account for some of

the inconsistent influence of the DTC measure on responsiveness to measured change over time. A negative DTC, as would be seen if the cueing effect were occurring, was present in about one third of the GABS sample at baseline for the DTC on gait velocity when counting backwards by ones. Other gait parameters and other secondary tasks that were considered for inclusion in the PI also demonstrated non-uniform effects of the secondary task on gait (i.e. a mix of participants with positive and negative DTC). A complementary explanation that may also account for some of this variability is that the current DTC paradigm is only concerned with gait performance and does not include in the score any indication for how well an individual performs the cognitive task. In spite of being instructed not to try harder on one task or the other, some people may prioritize the gait task while others prioritize the cognitive task, which could lead to different DTC scores even if brain capacity is similar.²³ Although this is more of a concern for individual-level than group-level measurement as the majority of people do demonstrate the expected gait worsening under dual-task conditions, revising the DTC paradigm to lessen some of these individual differences may improve group-level responsiveness of the GAB-PI.

7.2 Comparison with other ADAS-Cog Modifications

Scoring modification. First, the GAB-PI incorporates a statistical modification to scoring by standardizing component variables that originally had different scoring scales. Standardization allows these variables to be combined to produce a single final score with lower variability than the original raw score of any single component variable. One other ADAS-Cog 11 modification used a PI approach to combine the ADAS-Cog 11 with other cognitive tests, assessments of mood or behaviour, and assessments of the ability to perform activities of daily living. This PI demonstrated better responsiveness to a treatment effect in a clinical trial for Alzheimer's Disease than the ADAS-Cog 11 alone.³¹ In pre-dementia populations, three distinct approaches to scoring ADAS-Cog 13 (Section 3.3.1) data yielded different amounts of responsiveness to disease progression and responsiveness to treatment effects in MCI populations.³² In one instance, a re-weighted version of the ADAS-Cog 13 demonstrated greater responsiveness to disease progression in MCI than a version that included additional assessments of EF and functionality.³²

Taken together, this demonstrates that the way outcome measure data are analysed, and not just the content of that data, is important for outcome measure performance. The responsiveness of the GAB-PI may reflect a combination of the measures it comprises, and the PI approach used to combine them.

Additional cognitive assessments. In general, adding memory items to the ADAS-Cog 11, occasionally accompanied by the removal of other tasks, improved responsiveness to disease progression and to treatment effects.^{3,5,6,33} Adding assessments of EF to the ADAS-Cog 11, with or without modifying scoring methodology, and with or without additional measures of memory, was found to improve responsiveness to disease progression and to treatment effects in MCI populations in all but one instance.^{6,32-35} It is conceivable that at least some of the responsiveness of the GAB-PI to changes in pre-dementia syndromes is due to gait velocity and DTC picking up on EF or other cognitive abilities. For example, gait velocity has been associated with EF,^{15,18,22,25} and there is mixed evidence around the potential association of gait velocity or DTC gait velocity with memory.^{18,23,25}

Adding functionality assessments. It has been stated that ideal measures for MCI and early AD should include both cognitive and functional assessments,^{34,35} and the results of the literature review on ADAS-Cog 11 modifications support this. Modifications that added items to assess functionality, alone or in combination with other cognitive tests or alternative scoring methods, demonstrated superior responsiveness to disease progression and to treatment effects in MCI populations than ADAS-Cog 11 modification approaches that only modified cognitive content alone or in combination with scoring modification techniques.^{32-34,36} The only exception was that the ADAS-Tree outperformed the ADAS-Plus-EF&FA for responsiveness to disease progression, although the ADAS-Plus EF&FA demonstrated superior responsiveness over the ADAS-Tree to treatment effects³²; this point also serves to demonstrate the context specificity of responsiveness. The CDR-SB alone, which includes assessments of cognition and functionality, was also found to be more responsive to measured decline over two years in MCI and NC samples than several ADAS 11 modifications that re-weighted scores and/or added cognitive tests but did not include any assessment of functionality.⁵

Walking is a complex task,¹⁵ and gait velocity in particular has been considered a marker of functionality for older adults in the context of cognitive disorders specifically and ageing in general.^{15,16,30,37,38} Gait velocity and DTC are expected to reflect more subtle changes in functionality than those captured by the assessments used for previous ADAS-Cog modifications, such as the Functional Activities Questionnaire (FAQ), which rely on categorical response options of how well, if at all, a person can perform certain activities.³⁹ For example, the FAQ item assessing, through informant report, a subject's ability to perform "Shopping for clothes, household necessities, or groceries" may not capture changes that are noticeable and meaningful to a patient or caregiver but are not impactful enough to move from a categorical rating of "normal" to "has difficulty but does by self" or to "requires assistance". Nonetheless, during GAB-PI development it was explored whether adding an assessment of instrumental activities of daily living (ADL) to the PI would improve distinction between baseline diagnostic categories. This did not appear to be the case, so the simpler three item PI was retained (see Appendix B, Section B.2 for more information).

More generally, cognitive ability, the ability to move through one's environment, and the ability to perform mental and motor tasks at the same time, are all important aspects of functionality, and may be impacted by dementia and pre-dementia neuropathology. Even if a treatment does not alter the underlying pathology causing dementia, it can have a substantial impact on a patient's quality of life, as well as the burden held by caregivers, by stabilizing or slowing decline in cognitive ability, in motor function, or in the ability to maintain motor control while performing a cognitive task. There are several anecdotal reports of caregivers or clinicians noticing overall improvements in a patient when taking symptomatic treatments that are not accompanied by changes on standard tests of cognitive ability, and pharmacological treatments for dementia that demonstrate only modest effects on cognitive tests have been shown to delay admittance to nursing homes.⁴⁰⁻⁴² This suggests that cognitive tests alone are not capturing all important changes in disease severity. The GAB-PI was developed by selecting one measure to cover each of the three aspects of functionality presented above. Specifically, the ADAS-Cog-Proxy was selected to cover cognitive ability, single-task gait to cover movement, and dual-task gait cost to cover simultaneous cognitive and motor performance; however,

these three categories are not mutually exclusive. In particular, gait assessments may tap into cognitive domains, such as EF, that have been identified as important components of pre-dementia syndromes but are not covered by the ADAS-Cog 11.^{12,15,18,22,23,25,43} This introduces the idea that a single measure may serve the dual purpose of assessing both a manifest variable (gait velocity) and a latent variable (executive function), where both the manifest and latent variables are integral components of a disease, and changes in either the manifest or latent variable alone, or in combination, are meaningful. This contrasts with most cognitive assessments, including the ADAS-Cog 11, where the questions or tasks administered as manifest variable measures (e.g. the ability to draw a circle) are of much less importance than the latent variable that the resulting scores are intended to assess (e.g. praxis). The potential for gait assessments, under single-task or dual-task cost paradigms, to serve as a simultaneous measure of the fundamental bases for functionality—latent cognitive domain(s) and manifest motor function—may reduce inefficiencies in the measurement process and more closely represent whether treatments are having a meaningful impact on overall disease severity. Further research is needed to ascertain the validity of this statement, but the results of this thesis put into context with findings from other ADAS-Cog 11 modifications, suggest that gait assessments, supplemented with additional cognitive assessment, could play a valuable role in future studies of pre-dementia populations.

7.3 Limitations

The results of this thesis should be interpreted as preliminary, because a statistical model was used to approximate ADAS-Cog 11 scores in the GABS. More specific limitations to this process include slightly different study samples used to build the ADAS-Cog-Proxy from the GABS, and further error that may have been introduced through imputation of missing ADAS-Cog-Proxy covariates in the GABS. Furthermore, partly due to the nature of using data from an on-going cohort study, there were small sample sizes, especially at the longest follow-up time points. When assessing differences between people who did versus did not have each follow-up visit, it was found that participants with twenty-four and forty-eight month visits had statistically significant differences in gait speed (faster) than those who did not reach those follow-up visits. There were no statistically significant

differences in baseline gait velocity for the other lengths of follow-up, or for any follow-up length in age, education, DTC, or ADAS-Cog-Proxy scores.

When assessing responsiveness to disease progression, the standardized response mean (SRM) allows for direct comparison between the PI and the ADAS-Cog-Proxy, but the clinical meaningfulness of SRM units is not obvious. Furthermore, although this thesis only considered group-level responsiveness for the intended purpose of use in research studies, if individual-level responsiveness was of interest, such as for use in a clinical setting, a reference population would be needed to perform standardization of the PI component variables. It may be difficult to find a suitable pool of reference individuals.

In terms of the GAB-PI itself, the units are hard to interpret, and scoring calculations are more complex than for the ADAS-Cog 11. One major limitation is that the GAB-PI may not be useful for people with mobility impairment that is explainable by an event or disease other than dementia or pre-dementia syndromes (e.g. severe osteoarthritis of the lower limbs, Parkinson's Disease, amputation, stroke with residual motor deficits), even if some assessment of gait could be obtained. Most studies evaluating motor function in the context of dementia or pre-dementia syndromes, including the GABS, exclude individuals with severe mobility restrictions, which limits the generalizability of results.

On a broader level, based on the literature review in Chapter 3, the GAB-PI will be the thirty-second documented modified version of the ADAS-Cog 11. This thesis does not directly address the problem that has led to the creation of so many modifications: maintaining the ADAS-Cog 11 as the gold standard for assessing efficacy of treatments for dementia and pre-dementia populations, despite knowing it is not optimally responsive to important changes at the stages of disease where potential benefit of intervention may be greatest. While modifications can be beneficial in terms of within-study quality, they introduce a level of inconsistency and inefficiency that renders between-study comparisons, such as for meta-analyses, difficult; did intervention A find a benefit while intervention B did not because intervention A is truly more effective than intervention B, or because the ADAS-Cog 11 modification used to assess the efficacy of intervention A is more responsive to treatment effects than the one used to assess intervention B? Although modifications that maintain backwards compatibility with the

ADAS-Cog 11 have the potential to produce results on a standard metric, it is often up to the researchers of the study to produce these analyses, and the results are not necessarily expected to be consistent. By combining the ADAS-Cog 11 (or Proxy substitute) with quantitative gait assessments, the GAB-PI contributes to the heterogeneity of ADAS-Cog 11 modifications in a major way without a clear indication of how to reconcile results that may be inconsistent with those found on the ADAS-Cog 11 or other modified versions.

7.4 Suggestions for Further Research

Once an appropriate database with both the ADAS-Cog 11 and gait assessments becomes available, the PI can be re-built using the ADAS-Cog 11 instead of the ADAS-Cog-Proxy, and construct validation analyses comparing the GAB-PI to the ADAS-Cog 11 on responsiveness to baseline discrimination and to disease progression in a pre-dementia sample repeated. Test-retest as well as inter-rater reliability of the GAB-PI as a whole should be assessed, even though reliability of the individual components has previously been evaluated.^{44,45} If possible, it may also be of interest to compare the GAB-PI to other ADAS-Cog 11 modifications, especially those which incorporate assessments of functionality, and to repeat validation and reliability analyses for older adults with mild to severe dementia as an ideal outcome measure will be reliable, valid, and responsive to changes across the disease spectrum from NC to severe dementia.^{35,43,46}

The responsiveness of the GAB-PI to pharmacological and non-pharmacological treatment effects in pre-dementia populations should also be assessed. Although an outcome measure that is responsive to disease progression is often expected to also be responsive to treatment effects that slow, stop, or reverse disease progression, this is not always the case. For example, if a treatment targets one aspects of a disease that a given measure does not assess, that measure may miss important treatment effects but still be highly responsive to changes in natural history if it assesses other dynamic aspects of the disease that are not affected by the treatment. In the ADAS-Cog literature review most ADAS-Cog 11 modifications that demonstrated better responsiveness than the ADAS-Cog 11 to disease progression in pre-dementia populations also demonstrated superior responsiveness to treatment effects,^{3,5,32,33} but there was at least one notable exception to

this trend, which was a modification focused specifically on memory impairment.⁵ Thus, it will be important to assess the responsiveness of the GAB-PI to treatment effects in addition to disease progression. Ideally this would be done in the context of a treatment of known efficacy, but none exist for pre-dementia syndromes. Instead, the GAB-PI will be assessed as a secondary outcome measure for the SYNchronizing Exercises, Remedies in Gait and Cognition (SYNERGIC) clinical trial that will assess the effect of aerobic and progressive resistance training exercises, combined with cognitive training and vitamin D3, in older adults with MCI.⁴⁷ The SYNERGIC trial is currently recruiting participants in Ontario and British Columbia. The GAB-PI may be used to assess change over the 20-week study period within each treatment group (responsiveness to group-level within-person observed change over time), as well as between treatment groups (responsiveness to group-level between-person differences of within-person observed change over time), and compared to other outcome measures including the ADAS-Cog-Plus-EF&FA.

A final future direction revisits a key motivation behind this thesis, and proposes an alternative, more data-driven approach than ADAS-Cog 11 modification to explore the potential role of motor function assessments in detecting important changes in pre-dementia syndromes. The three components of the GAB-PI were selected based on theoretical, statistical, and practical considerations. Despite best efforts, these three components may not be the combination of measures most responsive to important changes in pre-dementia syndromes, such as those that occur between NC and MCI. A penalized regression analysis allows the simultaneous consideration of many more variables than can be sorted through manually.⁴⁸ This analysis will be used to obtain a model, composed of some subset of a pool of candidate variables, for discriminating between older adults with NC or SCI and older adults with MCI in GABS. The candidate pool of variables will contain all non-duplicate quantitative gait parameters collected by the electronic gait mat used in the GABS, including both those that have been studied substantially in the literature and those which have not, under single and the three dual-task conditions, as well as the DTC for each gait parameter and secondary task; balance assessments; the ADAS-Cog-Proxy and other global cognitive tests such as the Montreal Cognitive Assessment and domain specific cognitive tests; patient characteristics such as age and years of education; and ADL assessments. Given limits in human discernment,

this is far more variables than could be considered for inclusion in the GAB-PI, but an argument could be made for the potential utility of any of them for baseline discrimination purposes. The penalized regression analysis gradually provides the opportunity for more and more of these candidate variables to enter the model, and then cross validation can be used to obtain the best model, or regression equation, for baseline discrimination.⁴⁸ Further rationale for this analysis, alluded to earlier, is that the results of this thesis cannot discern if the addition of gait velocity and DTC to the ADAS-Cog 11 is leading to improvements in responsiveness beyond that which could be obtained by adding more standard assessments of functionality or additional cognitive tests; gait velocity may simply be serving as an assessment of traditional ADLs or EF rather than improving responsiveness because motor function is an integral component of the disease beyond that which may be captured by previously established outcome measures. Note that this would not negate any practical advantages of using gait assessment over more traditional outcome measures, such as brevity of assessment and language independence. Taking this one step further, from a purely measurement perspective, the ability to detect important changes is more important than the explanation about why these changes are being detected. The penalized regression analysis is a data-driven method that will help to identify whether motor function assessments (and if so, which ones) are selected for inclusion in a model for baseline discrimination between pre-dementia syndromes, even when the model has the option of global and domain specific cognitive tests, more general assessments of functionality, and patient characteristics previously shown to be relevant for pre-dementia syndromes. If motor assessments are selected this will further support their potential utility for detecting important changes in pre-dementia syndromes, but will not identify why this may be the case.

7.5 Clinical Implications

As clinical syndromes, dementia and pre-dementia stages of disease involve more than cognitive impairment; a key aspect of these syndromes is the impact that cognitive deficits have on functionality. Walking, or gait performance, is also important for the ability to perform several basic ADLs and has been postulated as a marker of overall functionality and a sixth vital sign to be assessed in older adults.^{16,49} Gait, and particularly

dual-task gait performance, has been shown to capture subtle changes during the cognitive decline associated with ageing and neurodegeneration, especially in MCI populations, which are not always captured by cognitive testing.²⁹ This thesis has shown that adding gait assessments to the ADAS-Cog 11 may help to differentiate between cognitively defined diagnostic groups, and detect progression in disease severity over time. The GAB-PI is also expected to be sensitive to changes in progression due to pharmacological and non-pharmacological treatments. To the knowledge of those who have contributed to this thesis, this is the first attempt to incorporate motor function assessments, which have previously been associated with progression to dementia,^{50,51} with the ADAS-Cog 11 which is the gold standard for testing interventions in dementia and pre-dementia syndromes.

7.6 Conclusion

This thesis has highlighted three challenges related to outcome measurement development and use, including those pertaining to the maintenance of a long-standing gold standard in a rapidly developing research field, to how outcome measures may be modified for improvement, and to the situation of performing secondary data analysis when not all necessary outcome measures are present in a single database. These challenges were considered in the field of dementia research. Overall this thesis suggests that the gold standard ADAS-Cog 11 is not an ideal outcome measure for studies on pre-dementia syndromes, and that improvements may be made by adding quantitative gait assessments to the ADAS-Cog 11 using a PI approach.

7.7 References

1. Grundman M, Petersen RC, Ferris SH, et al. Mild Cognitive Impairment can be distinguished from Alzheimer Disease and normal aging for clinical trials. *Arch Neurol*. 2004;61(1):59-66. doi:10.1001/archneur.61.1.59.
2. Labos E, Soderlund M, Trojanoski S, Seinhart D. Presentation of ADAS-Cog test with delayed recall and semantic encoding during the learning phase: results in normal, MCI and dementia population. *Alzheimer's Dement*. 2011;7(4):S250-S251. doi:10.1016/j.jalz.2011.05.713.
3. Sano M, Raman R, Emond J, et al. Adding delayed recall to the Alzheimer Disease Assessment Scale is useful in studies of Mild Cognitive Impairment but not Alzheimer Disease. *Alzheimer Dis Assoc Disord*. 2011;25(2):122-127.
4. Mahoney R, Johnston K, Katona C, Maxmin K, Livingston G. The TE4D-Cog: A new test for detecting early dementia in English-speaking populations. *Int J Geriatr Psychiatry*. 2005;20:1172-1179. doi:10.1002/gps.1412.
5. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav*. 2012;6:502-516. doi:10.1007/s11682-012-9186-z.
6. Podhorna J, Krahnke T, Shear M, E Harrison J. Alzheimer's Disease Assessment Scale-Cognitive Subscale variants in Mild Cognitive Impairment and mild Alzheimer's Disease: change over time and the effect of enrichment strategies. *Alzheimers Res Ther*. 2016;8(1):8. doi:10.1186/s13195-016-0170-5.
7. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical state for the staging of dementia. *Br J Psychiatry*. 1982;140:566-572.
8. Morris JC, Ernesto C, Schafer K, et al. Clinical dementia rating training and reliability in multicenter studies: the Alzheimer's Disease Cooperative Study experience. *Neurology*. 1997;48(6):1508-1510. doi:10.1212/WNL.48.6.1508.
9. Morris JC, Mckeel DW, Fulling K, Torack RM, Berg L. Validation of clinical diagnostic criteria for Alzheimer's Disease. *Ann Neurol*. 1988;24:17-22.
10. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*. 2010;74(3):201-209. doi:10.1212/WNL.0b013e3181cb3e25.
11. Buracchio T, Dodge HH, Howieson D, Wasserman D, Kaye J. The trajectory of gait speed preceding Mild Cognitive Impairment. *Arch Neurol*. 2010;67(8):980-986. doi:10.1001/archneurol.2010.159.
12. Montero-Odasso M, Verghese J, Beauchet O, Hausdorff JM. Gait and cognition: a complementary approach to understanding brain function and the risk of falling. *J Am Geriatr Soc*. 2012;60(11):2127-2136. doi:10.1111/j.1532-5415.2012.04209.x.
13. Lundin-olsson L, Nyberg L, Gustafson Y, Himbert D, Seknadji P, Karila-cohen D. "Stops walking when talking" as a predictor of falls in elderly people. *Lancet*. 1997;349:617.
14. Yogev-Seligmann G, Hausdorff JM, Giladi N. The role of executive function and attention in gait. *Mov Disord*. 2008;23(3):329-342. doi:10.1002/mds.21720.
15. Hausdorff JM, Yogev G, Springer S, Simon ES, Giladi N. Walking is more like catching than tapping: gait in the elderly as a complex cognitive task. *Exp Brain Res*. 2005;164(4):541-548. doi:10.1007/s00221-005-2280-3.
16. Studenski S, Perera S, Wallace D, et al. Physical performance measures in the

- clinical setting. *J Am Geriatr Soc.* 2003;51(3):314-322. doi:jgs51104.
17. Montero-Odasso M, Muir SW, Speechley M. Dual-task complexity affects gait in people with Mild Cognitive Impairment: the interplay between gait variability, dual tasking, and risk of falls. *Arch Phys Med Rehabil.* 2012;93(2):293-299. doi:10.1016/j.apmr.2011.08.026.
 18. Montero-Odasso M, Bergman H, Phillips NA, Wong CH, Sourial N, Chertkow H. Dual-tasking and gait in people with Mild Cognitive Impairment. The effect of working memory. *BMC Geriatr.* 2009;9(1):41. doi:10.1186/1471-2318-9-41.
 19. Allali G, Assal F, Kressig RW, Dubost V, Herrmann FR, Beauchet O. Impact of impaired executive function on gait stability. *Dement Geriatr Cogn Disord.* 2008;26(4):364-369. doi:10.1159/000162358.
 20. Hausdorff JM, Schweiger A, Herman T, Yogev-Seligmann G, Giladi N. Dual task decrements in gait among healthy older adults: contributing factors. *J Gerontol A Biol Sci Med Sci.* 2008;63(12):133501343. doi:10.1038/jid.2014.371.
 21. Hausdorff JM, Buchman AS. What links gait speed and MCI with dementia? A fresh look at the association between motor and cognitive function. *Journals Gerontol - Ser A Biol Sci Med Sci.* 2013;68(4):409-411. doi:10.1093/gerona/glt002.
 22. Rosano C, Studenski SA, Aizenstein HJ, Boudreau RM, Longstreth WT, Newman AB. Slower gait, slower information processing and smaller prefrontal area in older adults. *Age Ageing.* 2012;41(1):58-64. doi:10.1093/ageing/afr113.
 23. Nascimbeni A, Caruso S, Salatino A, et al. Dual task-related gait changes in patients with Mild Cognitive Impairment. *Funct Neurol.* 2015;30(1):59-65.
 24. Maquet D, Lekeu F, Warzee E, et al. Gait analysis in elderly adult patients with Mild Cognitive Impairment and patients with mild Alzheimer's Disease: simple versus dual task: a preliminary report. *Clin Physiol Funct Imaging.* 2010;30(1):51-56. doi:10.1111/j.1475-097X.2009.00903.x.
 25. Doi T, Shimada H, Makizako H, et al. Cognitive function and gait speed under normal and dual-task walking among older adults with Mild Cognitive Impairment. *BMC Neurol.* 2014;14(1):67. doi:10.1186/1471-2377-14-67.
 26. Gillain S, Warzee E, Lekeu F, et al. The value of instrumental gait analysis in elderly healthy, MCI or Alzheimer's Disease subjects and a comparison with other clinical tests used in single and dual-task conditions. *Ann Phys Rehabil Med.* 2009;52(6):453-474. doi:10.1016/j.rehab.2008.10.004.
 27. Muir SW, Speechley M, Wells J, Borrie M, Gopaul K, Montero-Odasso M. Gait assessment in Mild Cognitive Impairment and Alzheimer's Disease: the effect of dual-task challenges across the cognitive spectrum. *Gait Posture.* 2012;35(1):96-100. doi:10.1016/j.gaitpost.2011.08.014.
 28. Montero-Odasso M, Oteng-Amoako A, Speechley M, et al. The motor signature of Mild Cognitive Impairment: results from the Gait and Brain Study. *Journals Gerontol Ser A Biol Sci Med Sci.* 2014;69(11):1415-1421. doi:10.1093/gerona/glu155.
 29. Montero-Odasso MM, Sarquis-Adamson Y, Speechley M, et al. Association of dual-task gait with incident dementia in Mild Cognitive Impairment. *JAMA Neurol.* 2017;54(3):S154. doi:10.1001/jamaneurol.2017.0643.
 30. Auvinet B, Touzard C, Montestruc F, Delafond A, Goeb V. Gait disorders in the elderly and dual task gait analysis: a new approach for identifying motor

- phenotypes. *J Neuroeng Rehabil.* 2017;14(1):7. doi:10.1186/s12984-017-0218-1.
31. Carusone SC, Goldsmith CH, Smieja M, Loeb M. Summary measures were a useful alternative for analyzing therapeutic clinical trial data. *J Clin Epidemiol.* 2006;59(4):387-392. doi:10.1016/j.jclinepi.2005.05.009.
 32. Skinner J, Carvalho JO, Potter GG, Thames A, Zelinski E. The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): an expansion of the ADAS-Cog to improve responsiveness in MCI. *Brain Imaging Behav.* 2012;6:489-501. doi:10.1007/s11682-012-9166-3.
 33. Raghavan N, Samtani MN, Farnum M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimer's Dement.* 2013;9:1-19. doi:10.1016/j.jalz.2012.05.2187.
 34. Wessels A, Siemers E, Andersen S, et al. A combined measure of cognition and function for clinical trials: the integrated Alzheimer's Disease Rating Scale (iADRS). *J Prev Alzheimers Dis.* 2015;2(4):227-241. doi:10.1007/978-1-4614-5915-6.
 35. Huang Y, Ito K, Billing CB, Anziano RJ. Development of a straightforward and sensitive scale for MCI and early AD clinical trials. *Alzheimer's Dement.* 2015;11(4):404-414. doi:10.1016/j.jalz.2014.03.008.
 36. Wang J, Logovinsky V, Hendrix SB, et al. ADCOMS: a composite clinical outcome for prodromal Alzheimer's Disease trials. *J Neurol Neurosurg Psychiatry.* 2016;jnnp-2015-312383. doi:10.1136/jnnp-2015-312383.
 37. Lusardi MM, Pellecchia GL, Schulman M. Functional performance in community living older adults. *J Geriatr Phys Ther.* 2003;26(3):14-22.
 38. Adams K, Bayliss E, Blumenthal D, et al. Universal health outcome measures for older persons with multiple chronic conditions. *J Am Geriatr Soc.* 2012;60(12):2333-2341. doi:10.1111/j.1532-5415.2012.04240.x.
 39. Pfeffer R, Krosaki T, Harrah C, Chance J, Filos S. Measurement of functional activities in older adults in the community. *J Geront.* 1982;37(3):323-329.
 40. Lopez OL, Becker JT, Wisniewski S, Saxton J, Kaufer DI, Dekosky ST. Cholinesterase inhibitor treatment alters the natural history of Alzheimer's Disease. *J Neurol Neurosurg Psychiatry.* 2002;72:310-314. doi:10.1136/jnnp.72.3.310.
 41. Wallin A, Gustafson L, Sjogren M, Wattmo C, Minthon L. Five-Year outcome of cholinergic treatment of Alzheimer's Disease: early response predicts prolonged time until nursing home placement, but does not alter life expectancy. *Dement Geriatr Cogn Disord.* 2004;18:197-206.
 42. Beusterien KM, Thomas SK, Gause D, Kimel M, Arcona S, Mirski D. Impact of rivastigmine use on the risk of nursing home placement in a US sample. *CNS Drugs.* 2004;18(15):1143-1148. doi:10.2165/00023210-200418150-00008.
 43. Mohs RC, Knopman D, Petersen RC, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. *Alzheimer Dis Assoc Disor.* 1997;11:S13-S21.
 44. Montero-Odasso M, Casas A, Hansen KT, et al. Quantitative gait analysis under dual-task in older people with Mild Cognitive Impairment: a reliability study. *J Neuroeng Rehabil.* 2009;6(1):35. doi:10.1186/1743-0003-6-35.
 45. Grochowalski JH, Liu Y, Siedlecki KL. Examining the reliability of ADAS-Cog change scores. *Aging, Neuropsychol Cogn.* 2016;23(5).

46. Ferris SH. Cognitive outcome measures. *Alzheimer Dis Assoc Disord*. 1999;13:S140-S142.
47. Montero-Odasso M. SYNchronizing Exercises, Remedies in Gait and Cognition (SYNERGIC). <https://clinicaltrials.gov/ct2/show/NCT02808676>. Published 2017. Accessed July 2, 2017.
48. Hastie T, Qian J. Glnnet Vignett. https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html. Published 2014. Accessed January 2017.
49. Montero-Odasso M, Schapira M, Soriano ER, et al. Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older. *J Gerontol A Biol Sci Med Sci*. 2005;60(10):1304-1309. doi:10.1093/gerona/60.10.1304.
50. Kueper JK, Speechley M, Lingum NR, Montero-Odasso M. Motor function and incident dementia: a systematic review and meta-analysis. *Age Ageing*. 2017:1-10. doi:10.1093/ageing/afx084.
51. Beauchet O, Annweiler C, Callisaya ML, et al. Poor gait performance and prediction of dementia: results from a meta-analysis. *J Am Med Dir Assoc*. 2016. doi:10.1016/j.jamda.2015.12.092.

Appendices

Appendix A: Summary Table of ADAS-Cog 11 Modifications

| Outcome Measure | Test Items | Scoring Modifications | CV NC | CV MCI | CV Dementia |
|-------------------------------|--|---|---|--|--|
| ADAS-Cog 11 ¹ | Word recall, naming objects and fingers, commands, constructional praxis, ideational praxis, orientation, word recognition, language, comprehension of spoken language, word finding difficulty, and remembering test instructions | | 0.491 ² 0.468 ³ 0.468 ⁴ 0.452 ⁸ 0.468 ²⁵ | 0.383 ³ 0.383 ⁴ 0.381 ⁷ 0.302 ⁸ 0.412 ⁸ 0.452 ¹² 0.375 ¹⁹ 0.383 ²⁵ 0.384 ⁸⁶ | 0.627 ¹ 0.394 ² 0.335 ³ 0.339 ⁴ 0.517 ⁶ 0.390 ⁷ 0.333 ⁸ 0.320 ¹² 0.346 ¹³ 0.339 ¹⁹ 0.341 ²⁵ |
| ADAS-Cog 13 ⁷⁰ | ADAS-Cog 11, delayed recall, number cancellation or maze | | 0.432 ³ 0.447 ²⁵ | 0.333 ³ 0.439 ¹² 0.339 ²⁵ 0.337 ⁸⁶ | 0.263 ³ 0.249 ¹² 0.267 ²⁵ |
| VaDAS ^{72,73} | ADAS-Cog 11, two number cancellation tasks, delayed recall, maze, symbol digit modalities, digit backwards, animal category retrieval task | | N/A | N/A | N/A |
| ADAS-12 ^{7,17,75,76} | ADAS-Cog 11, delayed recall | | N/A | 0.343 ⁷ | 0.310 ⁷ |
| TE4D-Cog ⁶ | Commands, 7-word immediate recall, semantic memory (name seasons and match month to season), clock drawing, category fluency, orientation from ADAS-Cog 11, delayed recall | | N/A | N/A | 0.732 ⁶ |
| Pooled Index ⁷⁷ | Standardized ADAS-Cog 11, GDS, DBRI frequency, standardized MMSE, ADL, and DBRI | Pooled Index approach | N/A | N/A | N/A |
| ADAS-Rasch ⁷⁸ | ADAS-Cog 11 | Items weighted by measurement precision and based on IRT (OPLM) analysis. | 0.729 ³ 0.729 ²⁵ | 0.466 ³ 0.466 ²⁵ | 0.379 ³ 0.379 ²⁵ |
| ADAS-Tree ¹⁸ | ADAS-Cog 11, delayed word recall, number cancellation | Test items reweighted by random forest tree-based algorithm | 0.443 ³ 0.443 ²⁵ | 0.321 ³ 0.321 ²⁵ | 0.231 ³ 0.231 ²⁵ |
| cADAS-Cog ⁷⁹ | ADAS-Cog 11, delayed recall, number cancellation, maze | | N/A | N/A | N/A |

| | | | | | |
|----------------------------------|--|--|-----|-----|-----|
| | tasks; computerized multiphasic interactive neurocognitive display system for administration | | | | |
| TDAS ⁸⁰ | Word recognition, command, orientation, visual-spatial perception (modified constructional praxis), naming fingers, object recognition (modified naming objects), accuracy of the order of a process (modified ideational praxis), money calculation, clock time recognition (non-digital); computerized version of items administered with touch panel. | | N/A | N/A | N/A |
| CAMCOG-Plus ⁸¹ | Up to 25 items from CAMCOG (18 items in candidate pool), neuropsychological tests (12 items in candidate pool), ADAS-Cog (10 tasks in candidate pool), and MMSE (11 items in candidate pool) | Computer algorithm to estimate cognitive ability. Item difficulty estimated with OPLM and computerized algorithm selects next questions based on previous responses and continuously updates estimate of cognitive ability until 25 items administered or standard error of measurement for cognitive ability corresponds to 90% precision | N/A | N/A | N/A |
| ADAS-Cog -5 Subset ⁸² | Ideational praxis, remembering test instructions, language, comprehension, word finding difficulty | Subsetting analysis. Compare proportion of responders: responder=1=did not get worse over time on all items. Non-responder=0=subject got worse (pre-post diff >0) on at least one item | N/A | N/A | N/A |
| ADAS-Cog -6 Subset ⁸² | Word recall task, naming objects and fingers, commands, constructional praxis, orientation, and word recognition | Subsetting analysis: Responder=1=did not get worse over time on all items. Non-responder=0=subject | N/A | N/A | N/A |

| | | | | | |
|---|---|---|-----|----------------------|---------------------|
| | | ct got worse (pre-post diff >0) | | | |
| ADAS-bifactor ³ | ADAS-Cog 13 | Bi-factor model that accounts for residual correlations between word recognition and recall and between the 4 tasks based on interviewer report | N/A | N/A | N/A |
| ADAS-Cog-Plus-EF ³ | ADAS-Cog 13, category (vegetable) fluency | Bi-factor model | N/A | N/A | N/A |
| ADAS-Plus-EF&FA bifactor model ³ | ADAS-Cog 13, category (vegetable) fluency, TMT A&B, DSST, 5 Pfeffer FAQ items | Bi-factor model with methods factor for 5 FAQ items | N/A | N/A | N/A |
| Common Item Pooling ²⁰ | Various combinations from CAMCOG, ADAS-Cog, MMSE, neuropsychological tests | Common item pooling whereby OPLM used to estimate test item difficulty and subject cognitive ability, expressed as T-scores | N/A | N/A | N/A |
| ADNI Memory Composite ²⁵ | Word Recognition, RAVLT, delayed word recall, MMSE 3 word memory task with distractors, logical memory test | Factor score from single factor model | N/A | N/A | N/A |
| ADAS-Cog IRT ^{83,84} | ADAS-Cog 11 | IRT scoring methodology | N/A | N/A | N/A |
| ADAS-3 ¹⁹ | Word recall, delayed word recall, orientation | | N/A | N/A | N/A |
| CC1 ¹⁹ | ADAS-3, RAVLT Immediate, MMSE | Accounts for directionality of change | N/A | 10.930 ¹⁹ | 0.533 ¹⁹ |
| CC2 ¹⁹ | ADAS-3, cognitive portion CDR-SB | | N/A | 13.429 ¹⁹ | 0.538 ¹⁹ |
| CFC1 ¹⁹ | CC1, FAQ | | N/A | 9.273 ¹⁹ | 0.592 ¹⁹ |
| CFC2 ¹⁹ | CC2, FAQ | | N/A | 7.692 ¹⁹ | 0.609 ¹⁹ |
| Pharmacometric ADAS-Cog 13 ¹⁵ | ADAS-Cog 13 | Pharmacometric scoring methodology | N/A | N/A | N/A |
| IRT& Pharmacometric ADAS-Cog 13 ¹⁵ | ADAS-Cog 13 | IRT and Pharmacometric scoring methodology | N/A | N/A | N/A |

| | | | | | |
|---|---|--|-----|---------------------|---------------------|
| iADRS ⁸⁵ | ADAS-Cog 14, iADL OR substitute in ADAS-Cog 13, FAQ [ADAS-Cog 14 = ADAS-Cog 11, maze, digit cancellation, delayed recall] | $[-1(\text{ADAS-Cog14}) + 90] + \text{iADL}$ | N/A | N/A | N/A |
| Straightforward Sensitive Scale ⁸⁶ | Word recall, delayed word recall, orientation, FAQ, CDR-SB | | N/A | N/A | N/A |
| ADAS-Cog 3b ¹² | Word recall, orientation, word recognition | | N/A | 0.457 ¹² | 0.260 ¹² |
| ADAS-Cog 5 ¹² | ADAS-Cog 3b, delayed word recall, digit cancellation. | | N/A | 0.442 ¹² | 0.203 ¹² |
| ADAS-13RW ⁸⁷ | ADAS-Cog 13 | Reweight subsections as: memory = 0.10, clinician-rated tasks = 0.45, general cognitive tests = 0.45 | N/A | N/A | N/A |
| ADCOMS ⁸⁸ | Orientation, word recognition, word finding difficulty, delayed word recall, two MMSE items orientation time, drawing, and CDR-SB personal care, community affairs, home and hobbies, judgement and problem solving, memory, orientation) | Weighted linear combination based on partial least squares coefficients | N/A | 0.459 ⁸⁸ | N/A |

Abbreviations (in order of appearance): CV=Coefficient of Variation (Standard Deviation/Mean); NC=Normal Cognition; MCI=Mild Cognitive Impairment; ADAS-Cog=Alzheimer's Disease Assessment Scale-Cognitive Subscale; VaDAS=Vascular Dementia Assessment Scale; N/A=Not Available; GDS=Geriatric Depression Scale; DBRI=Dysfunctional Behaviour Rating Instrument; MMSE=Mini Mental State Examination, ADL=Activities of Daily Living; IRT=Item Response Theory; OPLM=One Parameter Logistic Model; cADAS-Cog=Computerized ADAS-Cog; TDAS=Touch Panel Type Dementia Assessment Scale ; CAMCOG=Cambridge Cognitive Examination; EF=Executive Function; FA=Functional Assessment; TMT=Trail Making Test; DSST=Digit Symbol Substitution Test; FAQ=Functional Assessment Questionnaire; ADNI=Alzheimer's Disease Neuroimaging Initiative; RAVLT=Rey Auditory Visual Learning Test; CC=Cognitive Composite; AVLT-Immed=Auditory Visual Learning Test-Immediate; CDR-SB=Clinical Dementia Rating Sum of Boxes; CFC=Cognitive Functional Composite; iADRS=Integrated Alzheimer's Disease Rating Scale; iADL=Instrumental Activities of Daily Living; ADAS-13RW=ADAS-Cog 13 Reweighted; ADCOMS=Alzheimer's Disease Composite Score. **Note:** Superscripts refer to reference numbers at the end of Chapter 3.

Appendix B: Supplementary Figures

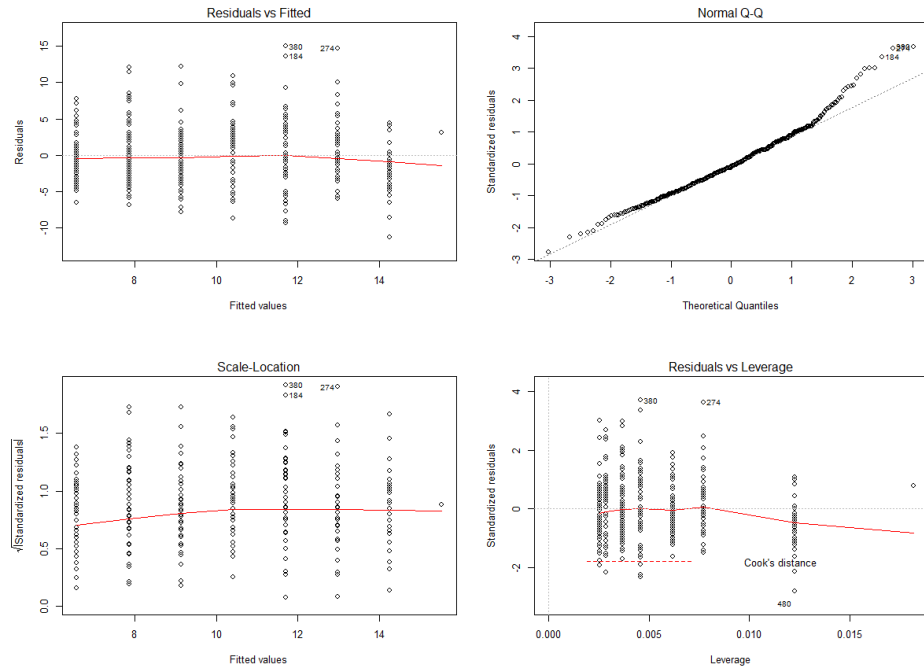


Figure B.1 Diagnostic plots for ADAS-Cog-Proxy Candidate Model 1.

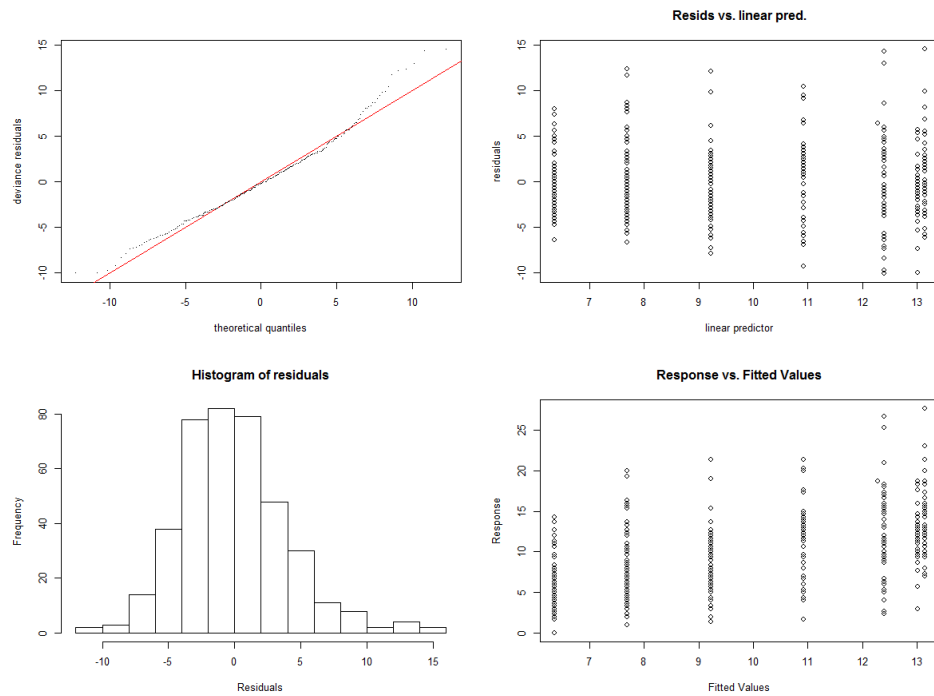


Figure B.2 Diagnostic plots for ADAS-Cog-Proxy Candidate Model 2.

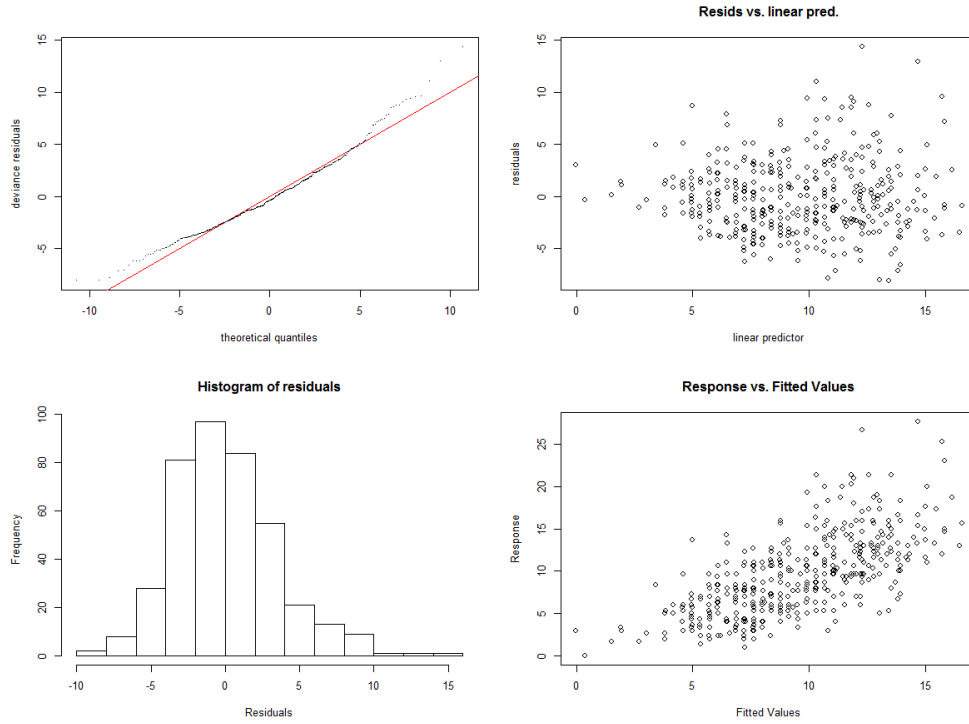


Figure B.3 Diagnostic plots for ADAS-Cog-Proxy Candidate Model 3.

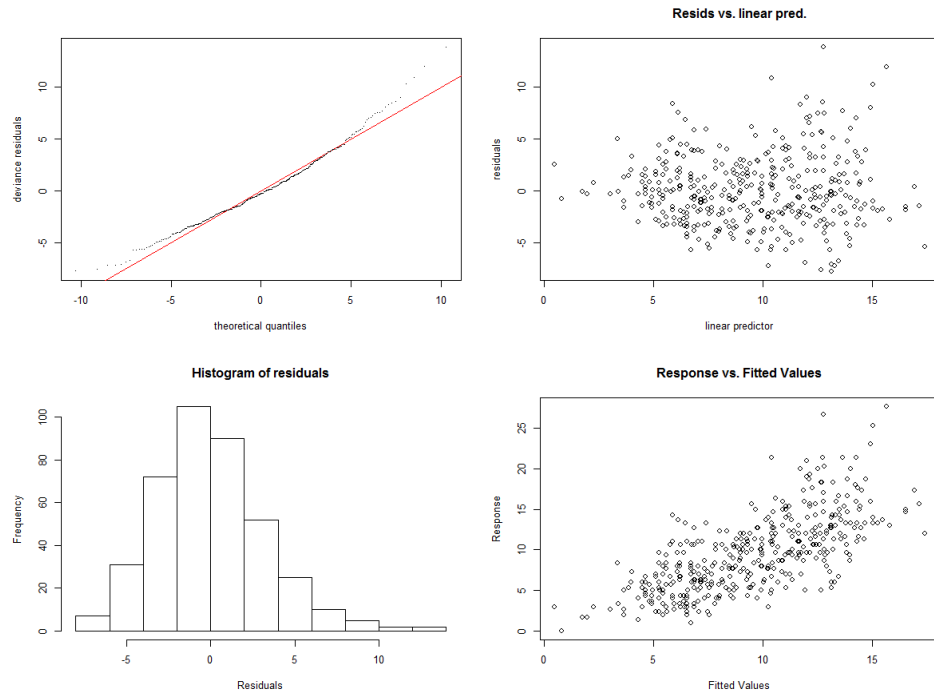


Figure B.4 Diagnostic plots for ADAS-Cog-Proxy Candidate Model 4.

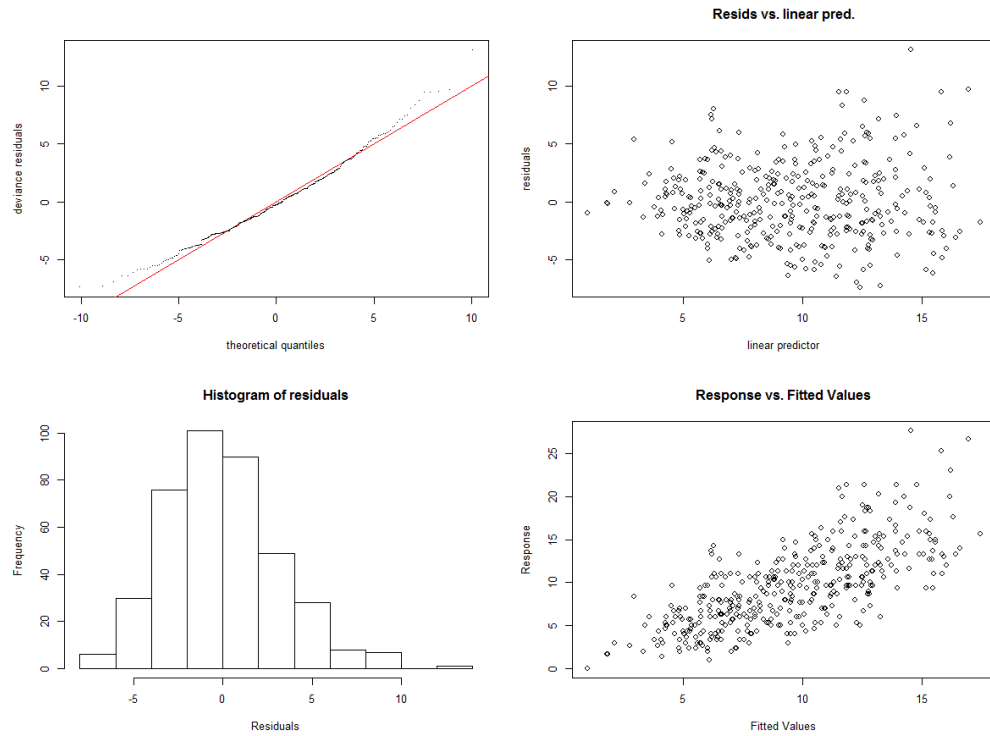


Figure B.5 Diagnostic plots for ADAS-Cog-Proxy Candidate Model 5.

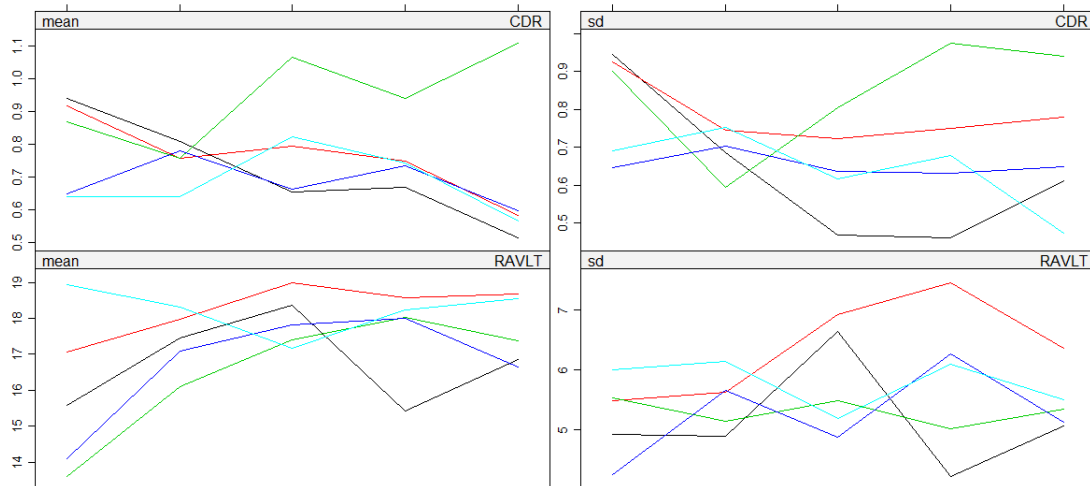


Figure B.6 Imputation streams for baseline assessment. Mean and standard deviation (SD) are for imputed values. Each line represents one imputation stream.

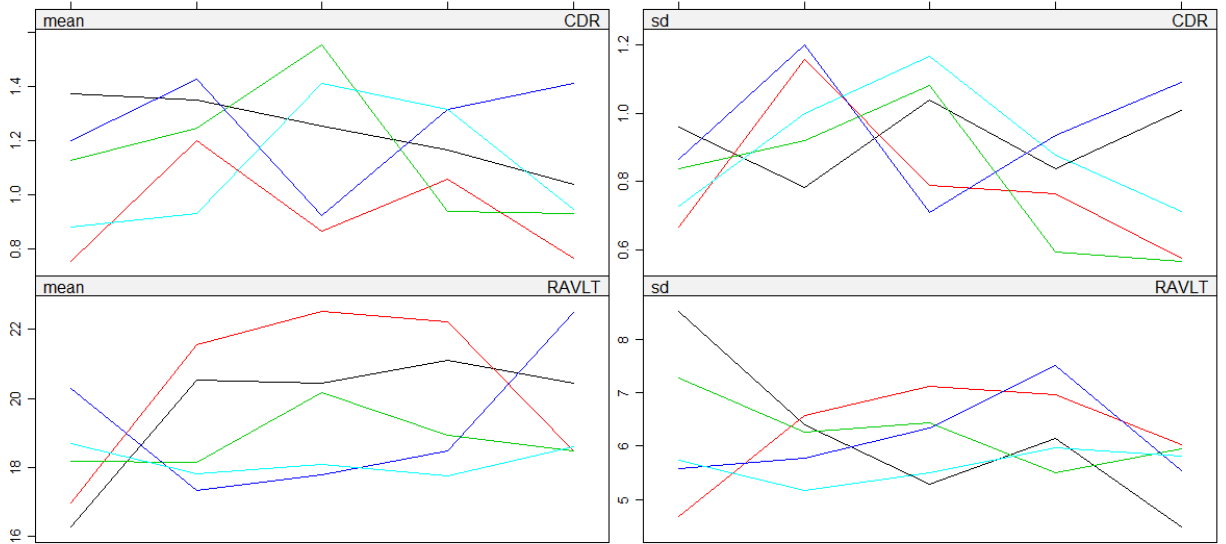


Figure B.7 Imputation streams for six-month follow-up assessment. Mean and standard deviation (SD) are for imputed values. Each line represents one imputation stream.

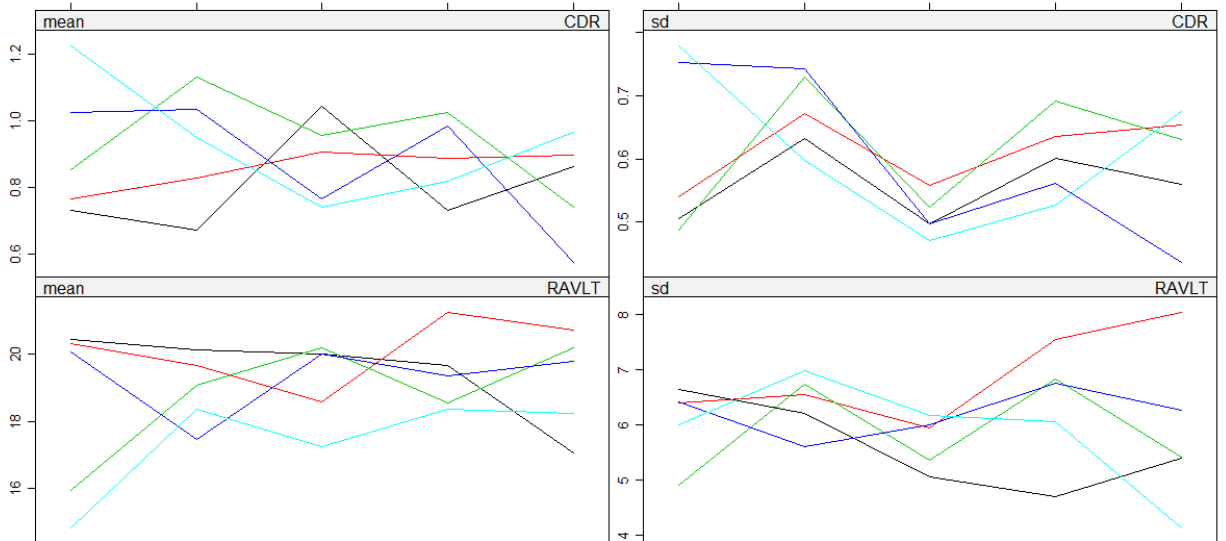


Figure B.8 Imputation streams for twelve-month follow-up assessment. Mean and standard deviation (SD) are for imputed values. Each line represents one imputation stream.

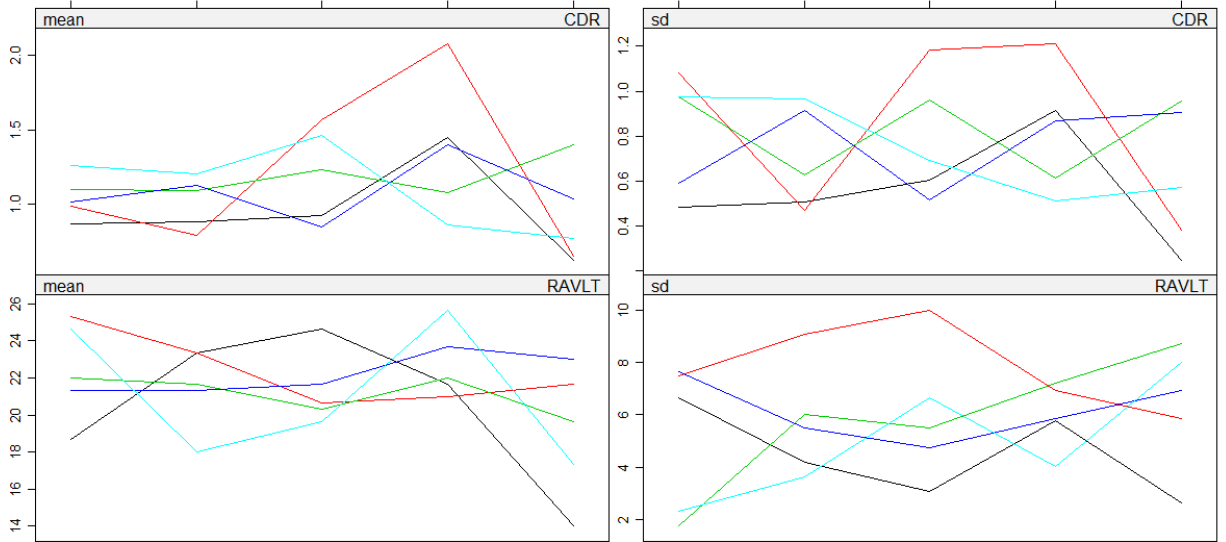


Figure B.9 Imputation streams for twenty-four-month follow-up assessment. Mean and standard deviation (SD) are for imputed values. Each line represents one imputation stream.

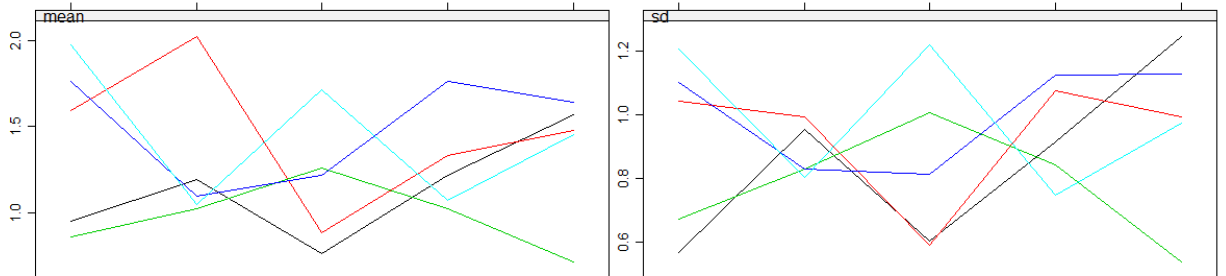


Figure B.10 Imputation streams for thirty-six-month follow-up assessment. Mean and standard deviation (SD) are for imputed CDR-SB values. Each line represents one imputation stream.

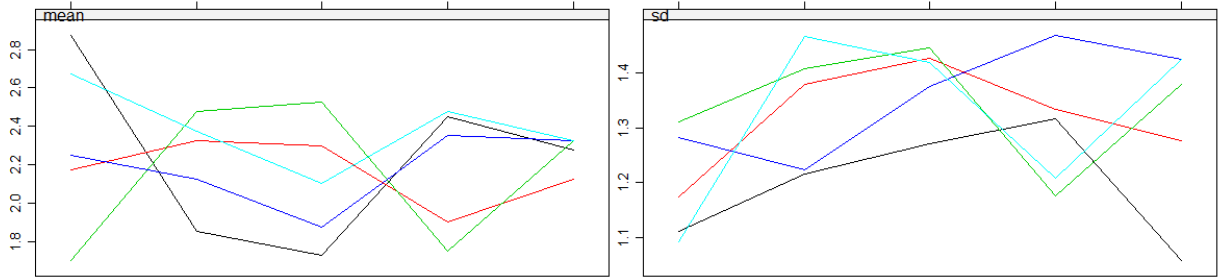


Figure B.11 Imputation streams for forty-eight-month follow-up assessment. Mean and standard deviation (SD) are for imputed CDR-SB values. Each line represents one imputation stream.

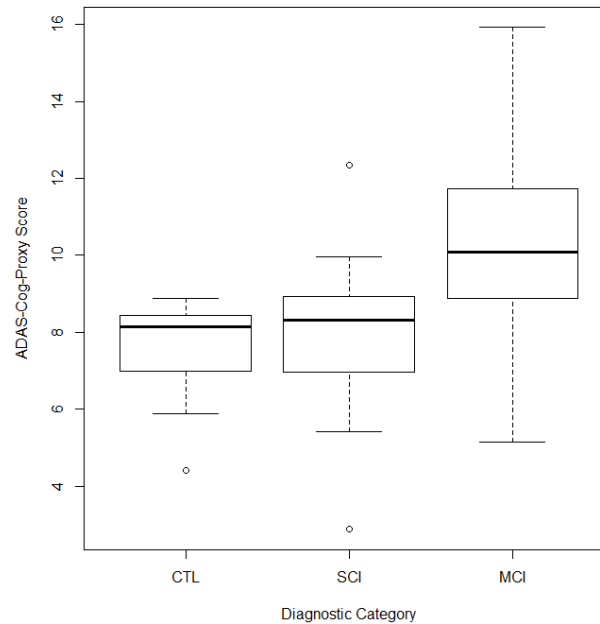


Figure B.12 Box plot summarizing distribution of ADAS-Cog-Proxy scores in each baseline diagnostic category (CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment).

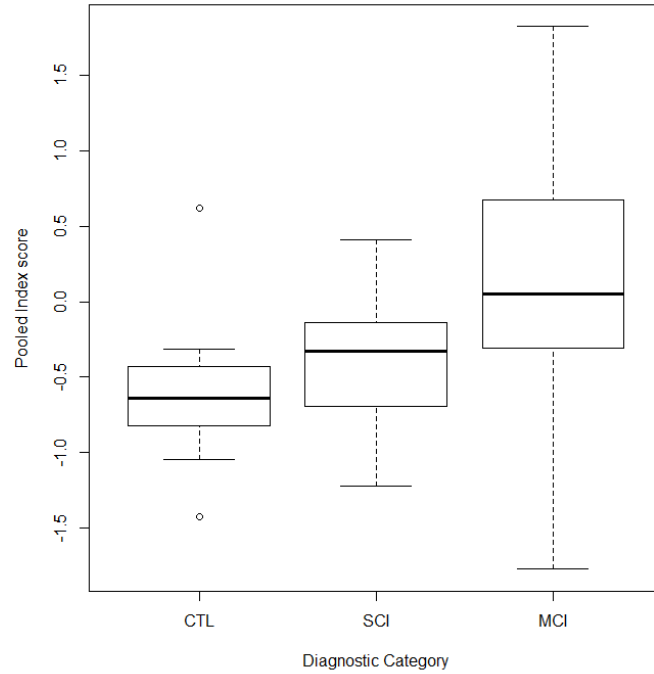


Figure B.13 Box plot summarizing distribution of Pooled Index scores in each baseline diagnostic category (CTL=Normal Cognition, SCI=Subjective Cognitive Impairment, MCI=Mild Cognitive Impairment).

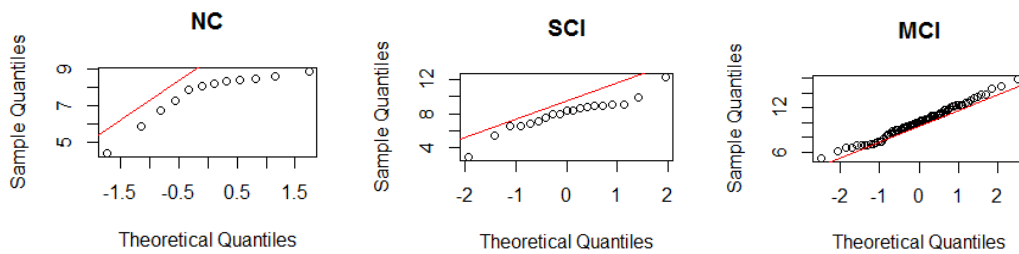


Figure B.14 QQ plots to assess normality of baseline ADAS-Cog-Proxy scores.

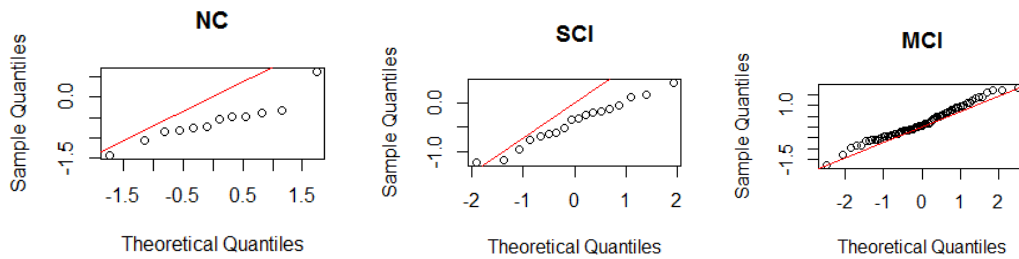


Figure B.15 QQ plots to assess normality of baseline Pooled Index scores.

Curriculum Vitae

Name: Jacqueline Kathleen Kueper

Post-secondary Education and Degrees: M.Sc. Epidemiology & Biostatistics (in-progress)
The University of Western Ontario
London, Ontario, Canada
2015-2017 (expected)

B.Sc. Biological, Biomedical, & Life Sciences
McGill University
Montreal, Quebec, Canada
2010-2015

Honours and Awards: Alzheimer Society of London and Middlesex Master's Scholarship
2016-2017

Canadian Society of Epidemiology and Biostatistics Travel Award
2017

Canadian Institutes of Health Research-Canadian Geriatrics Society Travel Award
2017

Dr. Carol Buck Graduate Scholarship in Epidemiology
2015-2016

McGill University J.W. McConnell Entrance Scholarship
2010-2011

Related Work Experience: Grader
Undergraduate Medical Education Program
Schulich School of Medicine and Dentistry
The University of Western Ontario
2016-2017

Graduate Teaching Assistant
Department of Epidemiology and Biostatistics
Schulich School of Medicine and Dentistry
The University of Western Ontario
2017

Publications:

Kueper JK, Speechley M, Lingum N, Montero-Odasso M. Motor function and incident dementia: a systematic review and meta-analysis. *Age and Ageing*. 2017:1-10. (Pre-Print) doi:10.1093/ageing/afx084.

Presentations:

Kueper JK, Lizotte DJ, Montero-Odasso M, Speechley M. Cognition and motor function: a novel outcome measure for studies of pre-dementia syndromes. Oral presentation at the *Canadian Society of Epidemiology and Biostatistics 2017 Conference: From Molecules to Population*, Banff, AB. Presented on 06/01/2017.

Kueper JK, Speechley M, Montero-Odasso M. One measure does not fit all. Is the ADAS-Cog responsive to important changes in pre-dementia studies? Poster presentation at the *Canadian Geriatrics Society 37th Annual Scientific Meeting*, Toronto, ON. Presented on 04/20-22/2017.

Kueper JK, Speechley M, Lingum N, Montero-Odasso M. Motor changes as a predictor of dementia syndromes: a systematic review. Poster presentation at the *Canadian Association on Gerontology 45th Annual Scientific & Educational Meeting*, Montreal, QC. Presented on: 08/21/2016.

Kueper JK, Speechley M, Lingum N, Montero-Odasso M. Motor changes as a predictor of dementia syndromes: a systematic review. Poster presentation at *London Health Research Day*, London, ON. Presented on: 03/29/16.