

Predicting dementia status from Mini-Mental State Exam scores using group-based trajectory
modelling

by

Cassandra Lynn Brown

B.A., Queen's University, 2008

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

© Cassandra Lynn Brown, 2012
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisory Committee

Predicting dementia status from Mini-Mental State Exam scores using group-based trajectory modeling

by

Cassandra Lynn Brown
B.A., Queen's University, 2008

Supervisory Committee

Dr. Andrea M. Piccinin, Department of Psychology
Supervisor

Dr. Holly Tuokko, Department of Psychology
Departmental Member

Abstract

Supervisory Committee

Dr. Andrea M. Piccinin, Department of Psychology

Supervisor

Dr. Holly Tuokko, Department of Psychology

Departmental Member

Background: Longitudinal studies enable the study of within person change over time in addition to between person differences. In longitudinal studies of older adult populations even when not the question of interest, identifying participants with dementia is desirable, and often necessary. Yet in practice, the time to collect information from each participant may be limited. Therefore some studies include only a brief general cognitive measure of which the Mini Mental State Examination (MMSE) is the most commonly used (Raina et al., 2009). The current study explores whether group-based trajectory modeling of MMSE scores with a selection of covariates can identify individuals who have or will develop dementia in an 8 year longitudinal study. Methods: The sample included 651 individuals from the Origins of Variance in the Oldest Old study of Swedish twins 80 years old or older (OCTO-Twin). Participants had completed the MMSE every two years, and cases of dementia were diagnosed according to DSM-III criteria. The accuracy of using the classes formed in growth mixture modeling and latent class growth modeling as indicative of dementia status was compared to that of more standard methods, the typical 24/30 cut score and a logistic regression. Results: A three-class quadratic model with covariate effects on class membership was found to best characterize the data. The classes were characterized as High Performing Late Decline, Rapidly Declining, and Decreasing Low Performance, and were labeled as such. Comparing the diagnostic accuracy of the latent trajectory groups against simple methods; the sensitivity of the final model was lower but it was the same or superior in specificity, positive predictive value, negative predictive value, and allowed a more fine-

grained analysis of participant risk. Conclusions: Group-based trajectory models may be helpful for grouping longitudinal study participants, particularly if sensitivity is not the primary concern.

Table of Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	v
List of Tables	vi
List of Figures	vii
Acknowledgments.....	viii
Introduction.....	1
Dementia	2
Longitudinal studies of aging.....	3
History of the Mini Mental State Examination.....	5
Sensitivity and specificity of the MMSE for dementia screening	7
Improvements in using the MMSE for dementia diagnosis.....	11
Tracking change in cognitive function using the MMSE	12
Group-based trajectory modeling.....	14
Current Study	17
Methods.....	19
Participants.....	19
Procedure	20
Measures	20
Demographic Information.....	20
Mini Mental State Exam (MMSE; Folstein et al.,1975).....	21
Cognitive Functioning.....	21
Dementia.....	21
Data Analysis	21
Sensitivity and specificity	21
Positive predictive value and negative predictive value	23
Logistic regression	23
Trajectory modeling.....	24
Results.....	29
Sensitivity and specificity	29
Logistic regression	30
Unconditional growth mixture models	30
Conditional models	34
Classification Accuracy	41
Posterior Probabilities	44
Analysis of subgroup with complete data until at least assessment wave 4	46
Discussion	51
References.....	64

List of Tables

Table 1. Demographic information and dementia status of participants.	29
Table 2. Cross-tabulation of classification based on formal diagnosis and on MMSE less than 24 proxy at first and fifth assessments.	30
Table 3. Fit indices for unconditional models.	32
Table 4. Parameter estimates of the unconditional quadratic models.	32
Table 5. Parameter estimates of the one-class conditional quadratic model.	35
Table 6. Fit indices for conditional quadratic models.	35
Table 7. Parameter estimates for the two and three-class conditional quadratic models.	36
Table 8. Covariate effects on the three-class quadratic model.	40
Table 9. Cross-tabulation of unconditional and conditional two-class model classification and dementia diagnostic status.	42
Table 10. Three-class unconditional quadratic model diagnosis of dementia and survival in study by wave.	44
Table 11. Three-class conditional quadratic model diagnosis of dementia and survival in study by wave.	44
Table 12. Fit indices for the unconditional LCGM of participants in study until at least wave 4.	48
Table 13. Fit indices for the conditional LCGM of participants in study until at least wave 4.	48
Table 14. Parameter estimates of three-class unconditional LGCM (restricted sample).	48
Table 15. Parameter estimates of the three-class conditional LGCM (restricted sample).	49
Table 16. Three-class unconditional quadratic latent class growth model with diagnosis of dementia at each wave by class (restricted sample).	49

List of Figures

Figure 1. A random selection of individual observed trajectories.	31
Figure 2. Average scores for the classes in the two-class unconditional quadratic model.	33
Figure 3. Average scores for the classes in the three-class unconditional quadratic model.	33
Figure 4. Average scores for the classes of the two-class conditional quadratic model.	37
Figure 5. Average scores for each of the classes in the three-class conditional quadratic model.	38
Figure 6. Observed trajectories of individuals in the "High performing Late Decline" class of the three-class conditional quadratic model.	38
Figure 7. Observed scores of individuals in the "Rapid Decline" class of the three-class conditional quadratic model.	39
Figure 8. Observed scores of individuals in the "Decreasing Low Performance" class of the three-class conditional quadratic model.	39

Acknowledgments

I would like to thank my supervisor, Dr. Andrea Piccinin for her continued guidance, encouragement, and support. The mentorship I have received has been invaluable to my development as a researcher. I would also like to acknowledge my clinical supervisor, Dr. Holly Tuokko for her expert feedback and advice. Further, I would like to thank Dr. Boo Johansson, and all the participants of the OCTO-Twin study without whom this work would not be possible. Finally, I'd like acknowledge the University of Victoria for their generous graduate fellowships in support of this research.

Introduction

Longitudinal studies have become an important resource for investigating the complexities of aging (Hofer & Sliwinski, 2001). There are currently over 100 completed or in progress longitudinal studies of aging around the world (Hofer & Piccinin, 2007). The results from such studies have challenged theories of aging based on cross-sectional studies (Hofer & Sliwinski, 2001). Dementias remain some of the most prevalent and concerning diseases that primarily affect older adults (Abbott, 2011). Large-scale longitudinal studies provide valuable information on dementia and its risk factors. For this reason, several longitudinal studies, such as the Canadian Study of Health and Aging, have aimed to investigate dementia (McDowell, Hill, & Lindsay, 2001). However, even in studies not primarily focussed on dementia, its prevalence in older populations means that having some knowledge of the cognitive status of participants is typically desirable. However, like all studies, the ability to collect information in longitudinal research is limited by the availability of participants and their willingness to continue answering questions and completing tests. This means that full diagnostic testing is frequently impossible and cognitive testing may be limited to short general cognitive measures, typically the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975). Yet, the collection of longitudinal data is expensive, labour intensive, and slow, so using the collected information to the fullest extent possible is a worthwhile endeavor. An effective method for identifying dementia cases could also be useful for future longitudinal studies. The current study investigates whether statistical methods can be used to improve the identification of cases of dementia using the most commonly used screening measure (Raina et al., 2009), the Mini-Mental State Examination (Folstein et al., 1975).

Dementia

Dementia is most commonly diagnosed based on the Diagnostic and Statistical Manual Criteria (DSM III or IV; American Psychiatric Association, 1987; American Psychiatric Association, 1994) or on those of the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer Disease and the Related Disorders Association (NINCDS-ADRDA; McKhann, Drachman, Folstein, Katzman, Price, & Stadlan, 1984; Mckann et al., 2011). Dementia is a diagnostic category that describes a particular group of symptoms. In both the DSM-IV and the NINCDS-ADRDA criteria, the diagnosis of dementia requires evidence of impairment in at least two domains of cognitive functioning. There must also be evidence of impairment in daily functioning and the impairment, cognitive and functional, must represent a decline from previous performance. Using the NINCDS-ADRDA criteria, “all cause dementia” is established first and then a more specific diagnosis of the etiology is made. Similarly, in the DSM-IV, dementia is considered the broader categorization and then specific diagnosis, such as Dementia of the Alzheimer’s Type, is made. Alzheimer’s disease is the most common type of dementia (Abbott, 2011; Alzheimer's Society of Canada, 2010). The older, DSM-III, diagnostic process involves first ensuring diagnostic criteria for dementia is met and then specifying the subtype if possible (American Psychiatric Association, 1987). According to the DSM-IV, all types of dementia are diagnosed as a dichotomy where an individual either meets criteria or does not. A diagnosis of Probable AD dementia suggests a greater level of certainty in the etiology than the diagnosis of Possible AD dementia, which can include less typical presentations and mixed etiologies (McKhann et al., 2011). The presence of dementia is diagnosed as a dichotomy, but the severity of dementia can be described based on the level of cognitive and functional impairment. The Clinical Dementia Rating Scale is one example of a scale used to quantify the severity or stage of cognitive and functional impairment in dementia (0 = normal, 0.5 = very mild dementia, 1 = mild dementia, 2 = moderate dementia, and 3 = severe dementia) (Morris, 1993).

Although the diagnostic subtypes of dementia are, in theory, based on different etiologies our understanding of what causes dementia remains incomplete. Current evidence suggests that a complex interplay of genetic, environmental, and lifestyle factors, such as diet education and activity, contribute to the development of dementia

There is a continued need for rigorous examination of hypothesized associations between lifestyle factors and dementia (Daviglius et al., 2011), particularly considering that even a 10-25% reduction in modifiable risk factors could potentially prevent 1-3 million cases of dementia worldwide (Barnes & Yaffe, 2011). Large sample sizes are often needed, particularly to study gene-environment interactions and modifiable risk factors (Eisenstein, 2011). More than large samples, teasing out antecedents from the consequences in the relations between lifestyle factors and dementia requires tracking individuals over time.

A decline in cognitive functioning is a requirement for diagnosing dementia, yet, changes in cognitive function are also part of the normative aging process. Declines are typically seen in speed of processing, working memory, inhibitory function, and long-term memory (Park & Reuter-Lorenz, 2009). Moreover, changes in cognitive function not due to dementia can also negatively impact functional capacity in older adults (Burton, Strauss, Bunce, Hunter, & Hultsch, 2009). As governments increase the age of retirement benefits, older individuals may be in the workforce longer, making understanding normative cognitive aging all the more pressing (Government of Canada, 2012). Accurate identification of dementia not only increases the opportunity for research focused on the disease, but also aids in isolating normative groups of older adults.

Longitudinal studies of aging

The National Institute on Aging and the Alzheimer's Association workgroup recommend in their update of the 1984 NINCDS-ADRDA criteria for dementia, that cognitive impairment be diagnosed through interviews with the patient and a knowledgeable informant, and an objective

cognitive assessment (McKann et al., 2011). They further specify that “bedside” mental status testing should be followed with neuropsychological testing when a confident diagnosis cannot be established from the patient’s history and mental status test results alone. Although ideal, neuropsychological testing and a review of patient history as would be done in a clinical setting is typically not practical in large longitudinal studies (Raina, et al., 2009). This issue is not unique to diagnosis of dementia, as there are many diagnoses of interest in longitudinal studies of older adults. Studies have addressed the issue in a variety of ways, including algorithms that incorporate multiple types of information (Raina, et al., 2009). Some longitudinal studies, such as the Origins of Variance in the Oldest-Old study (McClearn et al., 1997), conduct their own diagnostic assessments of participants for dementia. Some, including OCTO-Twin, identify the subtype of dementia, using a combination of neuropsychological measures and other information interpreted by physicians and health professionals similar to how it would be done in clinical practice. In the Canadian Study of Health and Aging, diagnosis was based on all available information by consensus of a nurse, physician, and neuropsychologist using DSM-III-R criteria (Tuokko, Morris, & Ebert, 2005). These likely represent the most reliable but time-consuming and labour intensive strategies for identifying participants with dementia or other diagnosable cognitive impairment.

Various attempts have been made to increase the efficiency of evaluating the dementia status of a large number of participants in longitudinal studies. For example, the Canberra Longitudinal Study used a standard protocol conducted by lay interviewers to target the information needed for DSM-III-R and ICD-10 diagnoses of dementia (Christensen et al., 2004). This strategy was a compromise: a select few tests and questions that target the domains of functioning typically tested by a larger battery, were used to determine a diagnosis (A. Mackinnon, personal communication, June 13, 2012). Others have attempted to maximize efficiency and include a broader range of participants by conducting some of the study protocol over the phone (Manly et al., 2011).

Longitudinal studies of aging that did not attempt to diagnose dementia frequently used general measures of cognitive function (Raina, et al., 2009). In a review of diagnostic tools and criteria used in longitudinal studies, Raina et al. (2009) found, out of 200 studies, the most commonly used measures were the MMSE, CAMCOG (Cognitive and self-contained part of the Cambridge Examination for Mental Disorders of the Elderly), CAMDEX (Cambridge Examination for Mental Disorders of the Elderly), Consortium to Establish a Registry for Alzheimer's Disease (CERAD), Geriatric Mental State (GMS), and the Automated Geriatric Examination for Computer Assisted Taxonomy (AGECAT). The criteria used in all 200 studies were DSM-III or IV, or NINCDS-ADRA. The most common measure was the MMSE, used in 140 out of the 200 studies included in the review (Raina, et al., 2009). The MMSE is intended as a screening tool rather than a diagnostic measure, but its sheer popularity as a measure used in longitudinal studies makes it an appealing candidate for attempts to improve the identification of cases of dementia. The current study evaluates whether trajectories of performance on the MMSE can be used to improve identification of dementia over simpler methods that do not include longitudinal information. First, the accuracy of more common methods, such as cut-off scores and logistic regression, will be reviewed and then tested in a sample where a diagnosis of dementia, where required, was made following a diagnostic assessment. Then, whether or not the trajectory of MMSE scores across multiple assessment occasions can be used to better predict dementia status will be examined using group-based trajectory modeling.

History of the Mini Mental State Examination

The MMSE was developed as a brief measure to quantify the cognitive status of hospital patients at bedside. It originated as a screening tool that could be administered quickly by clinicians and used to reliably differentiate patients with cognitive deficits from those without (Folstein et al., 1975). The items on the MMSE were included because they were already being used by clinicians to quickly assess the cognitive status of patients, and because their validity as indicators of cognitive

function would be readily apparent to non-health professionals (Folstein et al. 1975). This general cognitive measure includes 30 points addressing orientation to time and place, attention, recall, language, repetition, and comprehension. However, some areas of cognitive function are given more weight by virtue of having more items (points) testing that particular domain. The MMSE was originally tested on a group of 69 patients, 29 of which had dementia syndromes due to a variety of brain diseases, 10 had affective disorder, depressed type with cognitive impairment, and 30 had affective disorder, depressed type uncomplicated (Folstein et al. 1975). Patients were compared with a group of 63 “control” older adults who were not hospital patients. Patients with dementia had the lowest mean score; those with depression with cognitive impairment scored the second lowest on average, while those with depression without cognitive impairment scored on average only slightly lower than controls (Folstein et al. 1975). The MMSE was then standardized on a group of 137 hospital patients with a variety of conditions including affective disorder depressed type, affective disorder manic type, schizophrenia, personality disorder with drug abuse, neurosis, and dementia. The differentiation of patient groups by mean MMSE was confirmed.

Since the original validation study of the MMSE was published in 1975, the MMSE has become one of the most commonly used measures of cognitive function in research and practice (Nilsson, 2007; Raina et al, 2009). In a survey of neurologists in the United Kingdom, 91% of respondents reported using the MMSE in clinical practice, with 51% stating that they used it frequently (Davey & Jamieson, 2004). Nilsson (2007) suggests that Folstein et al.’s 1975 paper introducing the MMSE is the most cited paper in the health sciences, with the original paper receiving 19,721 citations between 1977 and 2006. Interestingly, a few citations in 1977 were the first, suggesting that it took a few years for the popularity of the MMSE to take off, and it has done so exponentially since then (Nilsson, 2007). Although originally a clinical measure, the number of citations attests to the fact that the MMSE is also now used as a brief measure of cognitive function in a large number of research

studies as well. The benefits of the MMSE for the researcher are similar to those for the clinician. Research protocols are often limited by the willingness of the participants to continue answering questions and thus efficient measures are preferred. It is also likely that the vast knowledge accumulated about the MMSE contributes to its continued use in research. For example, the MMSE is also used as an outcome measure in studies of the effects of various health conditions on cognition, for example, hemodialysis (Bossola et al., 2011), stroke (Toglia, Fitzgerald, O'Dell, Mastrogiovanni, & Lin, 2011), and (high and low) blood pressure (Guo, Fratiglioni, Winblad, & Viitanen, 1997; Molander, Gustafson, & Lovheim, 2010). More recently, the MMSE has been used to track cognitive changes over time in longitudinal studies of aging (Small, Viitanen, & Bäckman, 1997; Wilkosz et al., 2010). Given the multitude of ways the MMSE is used in clinical practice and research, there have been numerous investigations into the efficiency and validity of the MMSE and its uses in various populations (Mitchell, 2009; Tombaugh & McIntyre, 1992).

Sensitivity and specificity of the MMSE for dementia screening

The MMSE has been widely used to fill the need for a brief screening measure for dementia in older adults (Davey & Jamieson, 2004; Folstein et al., 1975; Grober, Hall, Lipton, & Teresi, 2008; Raina, et al., 2009). A screening measure is only as good as its ability to correctly identify cases and correctly rule out non-cases in the target population. By definition, though, a screening measure is not the “gold-standard” for diagnosis, and as such is not expected to have perfect accuracy. Screening tools are important for selecting individuals who would benefit from additional tests to determine if they have the condition of interest. They need to be more efficient, typically meaning faster and less expensive, than the more intensive testing and should therefore be more easily administered to large numbers of people. Screening tools are also used in large epidemiological studies to exclude people outside of the target population from the study or as a first step in identifying a target population with a particular disease. The MMSE fits these requirements as it is relatively brief and inexpensive to

administer. However, use of the MMSE as a screening measure has demanded investigation into its effectiveness in this task.

The effectiveness of screening measures is often summarized in calculations of sensitivity and specificity. Sensitivity is calculated as the total number of cases that are correctly identified using some predetermined criteria (in our case a cut-off score) as impaired, out of the total number of cases in the sample. Positive predictive value can also be calculated, which provides the certainty that a given individual, identified by the screening measure as having cognitive impairment, actually is impaired. Whether or not true cognitive impairment exists is determined by more reliable criteria, typically DSM-IV or NINCDS-ADRDA (Mitchell, 2009; Tombaugh & McIntyre, 1992). Specificity is the total number of non-cases correctly identified by the cut-off score out of the total number of non-cases in the sample. Additionally, negative predictive value can be calculated, which provides the certainty that a person identified by the screening measure as not having impairment is actually free of cognitive impairment. Tombaugh and McIntyre (1992) conducted a thorough review of the literature on the psychometric properties of the MMSE, including sensitivity and specificity. Some of the studies they reviewed investigated the possibility of using the MMSE to detect cognitive impairment in individuals suffering neurological diseases other than dementia and psychiatric illness. However, Tombaugh & McIntyre (1992) found that in about 70% of studies an MMSE score below 23 was associated with a diagnosis of dementia in at least 79% of cases. This suggests low MMSE scores are associated with dementia more than any other condition. Tombaugh and McIntyre (1992) also note that severity of cognitive impairment in the dementia groups was the main factor that differentiated studies that found high and low sensitivity: sensitivity ranged from 44% in an impaired group with a mean MMSE score of greater than 20 (Huff et al., 1987) to 100% when the mean MMSE score of the dementia sample was less than 15 (Folstein et al., 1975). More recent studies yield similar results. One study including only older adults with very mild dementia (excluding those with MMSE scores below 18 and corresponding

to a Clinical Dementia Rating score of 0 or 0.5) found 75% sensitivity and 90% specificity using the MMSE cut-off of 23 (Grober et al., 2008).

Mitchell's (2009) review of the accuracy of the MMSE in detecting found that the majority of studies used the less than 24 cut-off. They found, when pooling the results of pure community samples, sensitivity was 85.1%, specificity was 85.5%, positive predictive value was 34.5%, and negative predictive value was 98.5%. Including primary care samples, the values were 78.4%, 87.8%, 53.65, and 95.7% respectively. In specialist settings, Mitchell found lower sensitivity (62.7%) and concluded that the majority of positive results on the MMSE would actually be false positives. It is interesting that the specialist setting, in which one might expect higher rates of dementia, and therefore better positive predictive value actually showed lower positive predictive values. It is possible, though, that individuals seeking specialist care are cases of very mild dementia, which are difficult to detect using gross screening measures. Mitchell's review included only the English version of the MMSE, but similar results have been found for the Swedish version using the same less than 24 cut-off: sensitivity was 87%, specificity 92% and positive predictive value 69% (Grut, Fratiglioni, Miitanen, & Winblad, 1993). Although the less than 24 cut-off identifies cognitively impaired versus healthy individuals fairly well for more severe cases, it is much poorer at identifying very mild dementia.

Another way to use the MMSE is to compare a person's results to the distribution of scores in the population, thus treating it more as a norm referenced test. Several studies have explored normative performance on the MMSE for a variety of populations including younger adults (Crum, Anthony, Bassett, & Folstein, 1993) older adults (Dufouil et al., 2000), and people with Alzheimer's disease (Rasmusson, Carson, Brookmeyer, Kawas, & Brandt, 1996). Such epidemiological studies have produced normative data including percentile rankings stratified by age, education, or a combination of the two for adults in the United States (Crum et al., 1993), the United Kingdom (Dufouil, et al., 2000), Brazil (Moraes, Pinto, Lopes, Litvoc, & Bottino, 2010), Sweden (Grut et al., 1993), and Australia

(Anderson, Sachdev, Brodaty, Trollor, & Andrews, 2007). Normative data are useful because it allows for a quick evaluation of whether a given score is particularly aberrant given some basic information such as how old the individual is, where they are from, and how many years of education they completed. While useful, one difficulty with normative data for diagnostic purposes is that the normative sample is often recruited without differentiating between healthy individuals and those with cognitive impairment due to other causes or other health issues that may interfere with test performance (Crum, et al., 1993; Dufouil, et al., 2000). The diagnosis of dementia also requires evidence that the level of cognitive function represents a decline from previous levels of ability (McKhann et al., 2011). Normative data cannot address, at the individual level, whether a decline from previous levels of functioning has occurred.

Normative studies of MMSE performance have found that age is related to scores (Anderson, et al., 2007; Crum, et al., 1993; Dufouil, et al., 2000; Moraes, et al., 2010). The correlation between age and MMSE scores has been reported as moderately negative (Crum et al. 1993). Dufouil et al. (2000) who provided percentiles based on age, sex, and education, found that differences were greater between levels of education in older age groups. For example, the 90th percentile of women who left school after they turned 15 was 30 at 75 years old and 28 at 95 years old. For women who left school before age 15 the 90th percentile was 29 at age 75 and 23 at age 95 (Dufouil, et al., 2000). The prevalence of cognitive impairment not due to dementia is known to increase with advancing age, which may be reflected in the average MMSE scores, alternatively individuals with dementia may be being included in normative samples. Cummings (1993) suggests that the MMSE items generally test knowledge accumulated over time, which is typically better preserved in normative aging, making it less likely that the lower average MMSE score of older populations is due to the effects of normative aging. When older adults scoring below the typical 24 cut-off were excluded from the analysis, the relationship between older age and lower MMSE scores persisted, suggesting that there is potentially an effect of

normative aging (Anderson et al. 2007). However, excluding older adults with an MMSE score below 24 does not guarantee that the remaining group is dementia free and so does not completely separate the two effects (Shiroky et al., 2007).

It may be through other variables such as the level of intellectual activity required by ones occupation, that level of education is related to MMSE scores (Anderson, et al., 2007). It may also be through general intelligence, or socioeconomic status (Anderson, et al. 2007). Brain reserve, a property of the physiology of the brain related to its ability to withstand insults and still function appropriately, may also be related to the range of MMSE scores in older adults. Reynolds, Johnston, Dodge, DeKosky, & Ganguli (1999) found that head circumference, used as a proxy for brain reserve, was related to MMSE scores such that for every 1 centimetre increase in head circumference there was a corresponding 20% reduction in the probability of being in the lowest scoring MMSE group. Interestingly, all of these factors, education, age, socioeconomic status, and brain reserve are related to the prevalence of diagnosed dementia. Thus it becomes difficult to disentangle whether lower MMSE scores in older, lower education groups, reflect greater incidence of true cognitive impairment, or whether the MMSE is biased in these groups. The test is biased if it consistently identifies certain groups as having cognitive impairment when they have none and vice versa. Using existing methods of screening with the MMSE, percentiles or cut-off scores, early dementia in high functioning individuals is likely to go unnoticed while lower functioning individuals will disproportionately screen positive.

Improvements in using the MMSE for dementia diagnosis

Several attempts to improve the accuracy of the MMSE in screening individuals with dementia have been made. These include adding questions (Cacho, Benito-Leon et al., 2010; Tombaugh, 2005), adapting questions to improve the accuracy for less educated older adults (Olazaran et al., 2004), and using a subset of items that might better predict dementia (Olazaran et al., 2004). Most attempts were focussed on improving the accuracy of detecting mild dementia, as it is for this group that the

sensitivity of the MMSE is poorest (Tombaugh & McIntyre, 1992). Cacho et al. (2010) found that, while using the 23/24 cut-off, adding the clock test to the MMSE increased the specificity from 86.4% to 89.4% in a sample of older adults with mild dementia. In the clock test the individual is asked to draw the face of a clock with the hands pointing to a particular time and the drawing is scored based on accuracy. A version of the MMSE was adapted for less educated older adults by adding some less demanding items to the original MMSE (Olazaran et al., 2004). In the adapted MMSE, digits backward was added to serial 7's, the instruction that is usually written was changed to a visual order of a man raising his arms, spelling 'world' backwards was not administered, and two overlapping circles were added to the design copy task.

There is some evidence that particular items on the MMSE are predictive of dementia. Olazaran et al. (2004) found that failure on the delayed-recall items indicated increased risk of dementia such that for every word correctly recalled the expected risk of future dementia decreased approximately two fold. Similarly, Brungnolo et al. (2009) found, in their two-factor solution for the MMSE in Alzheimer disease patients, the first factor scores decreased linearly with MMSE total scores whereas the second factor only decreased in patients with a total score of less than 21. They suggest the first factor, composed of orientation to time and place, delayed recall, attention/concentration, constructional praxis, and comprehension, may characterize a working memory factor that is sensitive to the effects of mild dementia. The ability to detect mild dementia using the MMSE may be improved by weighing items in the working memory factor more heavily than others, such as naming.

Tracking change in cognitive function using the MMSE

In their initial study Folstein et al. (1975) examined the test-retest reliability of the MMSE. They re-administered the test after 24 hours to a group of 22 patients with depressive symptoms and found a Pearson correlation of 0.89 between the two measures and no significant difference in performance. They also administered the MMSE again after 28 days to a group of 24

“clinically stable” patients with dementia, depression, and schizophrenia, and found an even higher Pearson correlation of 0.99, and a non-significant difference in mean MMSE scores across the two occasions. Folstein et al. (1975) suggested that these findings indicate the MMSE shows few practice effects and is suitable for repeated uses over time.

Several longitudinal studies have found that older adults who remain dementia free typically have relatively stable MMSE scores over time (Jacqmin-Gadda, Fabrigoule, Commenges, & Dartigues, 1997; Ratcliff et al., 2003). This is in contrast to cross sectional studies that have consistently shown a negative relationship between MMSE scores and age (Moraes et al., 2010; Anderson et al., 2007; Crum et al., 1993; Dufouil et al., 2000). Jacqmin-Gadda et al. (1997) found that the regression coefficient for age was different from that for time in their longitudinal study of older French adults, suggesting that there was either a cohort or practice effect, or both. Despite these studies finding little practice effects overall there are reports by clinicians of people “studying” for the MMSE (Tombaugh & McIntyre, 1992). The issue of practice effects plagues all tests that are administered more than once and has yet to be resolved in a satisfactory manner. Multiple forms can be used in some cases but ensuring the equivalency of forms can be challenging, and complicate interpretations.

Reliable change indices account for the reliability of the measure in an attempt to establish the size of change needed between two occasions of measurement for it to be considered true change rather than due to the unreliability of the measure. Tombaugh (2005) used reliable change indices in a sample of cognitively normal CSHA participants over 65 years old to establish the size of the difference between two MMSE scores that would be indicative of actual change. Participants completed the MMSE twice at the first occasion, then twice again five years later. Tombaugh (2005) found, using reliable change index regression and reliable change index difference scores, that the difference between MMSE scores at separate occasions needed to be greater than 5 points to be statistically significant. This suggests smaller differences may be due to the unreliability of the MMSE.

Interestingly, the sample means showed only a 0.5 point difference between administrations. This suggests that although mean MMSE scores for a healthy population appear stable, the scores of any given individual may be more variable.

Despite concerns about the lack of stability of MMSE scores even in healthy populations, the MMSE has been used as an outcome measure in studies of cognitive function, particularly in Alzheimer disease (Ratcliff et al., 2003; Wilkosz et al., 2010). However, not everyone with dementia shows the same pattern of cognitive decline, and other factors such as psychotic symptoms, lower baseline cognitive ability, and younger age at diagnosis appear to predict a more rapid course of decline (Davidson et al., 2010; Wilkosz et al., 2010).

The MMSE is used frequently in longitudinal studies, but as has been reviewed, common strategies for using MMSE scores to determine whether a particular participant has developed dementia or not, such as scoring less than 24, are of limited accuracy (Raina et al., 2009). We aim to use trajectories of performance on the MMSE, rather than single occasion data, to better predict the dementia status of older adults. Similar attempts to improve the diagnostic accuracy of the MMSE have been made. Small et al. (1997) used logistic regression to predict the categories cognitively intact or Alzheimer disease from MMSE scores. However, their aim was to explore whether specific items on the MMSE measured over time could predict Alzheimer disease. Wilkosz et al. (2010) used latent class trajectory models with the MMSE as an outcome measure, but their aim was to explore the trajectories of cognitive decline in Alzheimer disease and thus included only individuals with Possible or Probable Alzheimer disease already diagnosed.

Group-based trajectory modeling

Trajectories of within person change are frequently modeled in psychological literature using multilevel linear models or latent growth curve model approaches (Singer & Willett, 2003). In a multilevel modeling framework, within person change is described as a function of time and between

person differences are described by random effects and coefficients. From a structural equation modeling approach individual growth curves are considered latent variables with a covariance and mean structure (Muthén, 2004). In traditional latent growth modeling, the average growth trajectory is defined for the whole sample and deviance from the average is described as variance. Latent growth curve models assume that all individuals are drawn from the same population and that variability in growth trajectories can potentially be explained by covariates (Muthén, 2004).

Group-based trajectory modeling is essentially an extension of latent growth curve models such that individuals are assumed to be members of a finite number of subpopulations, rather than a single homogenous population (Muthén, 2004). Nagin & Tremblay (2005) refer to both latent class growth modeling and growth mixture modeling as group-based trajectory modeling. In all group-based trajectory modeling approaches, different subpopulations are characterized as belonging to latent classes, and the parameters that define the shape of the trajectory and the probability of class membership for each case are estimated. Thus, rather than individual differences characterized by variation around a single growth parameter mean, growth mixture models account for some of this variation by allowing growth parameters to vary around a different mean for each class specified within the sample (Muthén, 2004). Latent class growth models, which allow no variance around the different trajectories, attribute all trajectory differences to class. Growth mixture modeling is a mix of a person-centered approach, latent class membership, and a variable centered approach that allows the impact of covariates to differ by class (Jung & Wickrama, 2008).

Growth mixture modeling is a relatively new statistical approach, with the first uses appearing about fifteen years ago (Muthén & Shedden, 1999). Similar approaches modeling trajectories of subpopulations were first advanced to examine career criminality (Nagin, 1999; Nagin & Land, 1993). However, the usefulness of group-based trajectory modeling approaches extends to many cases where different developmental trajectories are expected. Such approaches allow the antecedents and

consequences of particular trajectories to be explored, which can be helpful in characterizing developmental risk factors and potential avenues of intervention. The early use of similar techniques to identify developmental trajectories of aggressive, antisocial and criminal behavior is one example (Schaeffer, Petras, Ialongo, Poduska, & Kellam, 2003). Other studies have used the approach to examine trajectories of adolescent alcohol use, including predictors and outcome such as alcohol related problems (Colder, Campbell, Ruel, Richardson, & Flay, 2002; Muthén & Shedden, 1999). Trajectory classes can also be helpful in treatment studies where different subpopulations may respond differently to intervention (Leoutsakos, Muthén, Breitner, Lyketsos, & Team., 2012). A particular trajectory group may also have greater or lesser likelihood of some outcome of interest such as dropping out of high school (Muthén & Asparoutiov, 2008). Such outcomes can be included directly in the model specification or can be explored in relation to classes identified without including outcome information (Petras & Masyn, 2010).

Growth mixture modeling is not without criticism. One is the difficulty of characterizing latent trajectory classes. Nagin and Trembly (2005) suggest that identifying classes may be a succinct way to characterize individual differences, but that classes may not necessarily represent taxonomies. Others argue that group-based trajectory models are only appropriate when true developmental typologies exist (Sampson, Laub, & Eggleston, 2004). Further criticisms include the subjectivity of enumerating the “correct” number of classes and the possibility of over specification (identifying more classes than truly exist because of between person differences) (Bauer & Curran, 2003). Further, the majority of previous studies using the growth mixture model approach to data analysis do so to explore potential classes (Muthén et al., 2002). Our approach is less common, in attempting to evaluate classification accuracy by a known criterion (Tofighi & Enders, 2008).

Small and Bäckman (2007) developed a two class model for an older adult sample that included individuals who would be, but had not yet, been diagnosed with dementia, as well as controls who

remained dementia free over the follow-up period. They compared the model classification to the known outcome of whether the person would be diagnosed with dementia at the last assessment to examine the accuracy of the model in identifying preclinical dementia cases. They did not include any covariates aside from age in the model. They did, however, examine the characteristics of individuals who were correctly classified by the model and those who were misclassified. They did not find any significant effects of gender. They found that those who would be diagnosed with dementia, regardless of the way they were classified by the model, were likely to be older. Those who were misclassified as preclinical dementia (false positives) had lower levels of education than those who were correctly identified as preclinical dementia (true positives) and those who were falsely identified as normal aging when they would actually be diagnosed with dementia (false negatives). This suggests that including covariates in the model specification may improve the model's classification accuracy. They also only considered the two class model; they did not explore whether a greater number of classes may have better represented the data. The current study explores this possibility.

Current Study

In the current study we aim to explore whether using group-based trajectory modeling techniques can improve the identification of dementia based on the MMSE in large longitudinal datasets. As a first step in illuminating whether group-based trajectory models can be used to this end, we develop a model first using a data set where diagnosis was made according to DSM-III-R criteria following a full assessment and case conference as the standard against which we compare the accuracy of the model classification, similar to the method used by Small and Bäckman (2007). The current analysis includes individuals who were diagnosed with dementia at baseline, those who develop dementia over the study period, and those who remain dementia free, unlike Small and Bäckman (2007) who include only individuals diagnosed at their last assessment. The current analysis will further expand on Small and Bäckman's (2007) findings in important ways. First, we explore whether

the accuracy can be improved by including the covariates sex, age, and years of education, in the model specification. We will also explore whether logistic regression, which uses only a single measurement occasion but includes covariates improves case identification. Then, we explore whether a model with more classes better represents the data. In short, the present study will compare the sensitivity, specificity, positive predictive value, and negative predictive value of the group-based trajectory model to the traditional 24 cut-off, and to a logistic regression analysis.

Methods

Participants

Participants were drawn from a population-based, longitudinal study the Origins of Variance in the Old-Old (OCTO Twin) Study (McClean et al., 1997). Potential participants were identified from the oldest cohort of the Swedish Twin Registry, which records the birth of all twins in the country. OCTO Twin participants were twin pairs born in 1913 or earlier where both twins were alive and either 80 years old or turning 80 within the 3 years in which the first wave of data was collected. Data collection began in 1991 with 737 pairs (1474 individuals), however, some of these pairs were excluded because one or both twin partners became deceased before they were scheduled for the first examination (188 pairs). A group of 198 pairs were excluded because one or both twins declined to participate in the study. A total of 702 individuals from 351 complete twin pairs participated in wave 1 of the study. Participants were assessed for the first time between 1991 and 1993, then every 2 years for a maximum of 5 times. Date and cause of death were available from external sources and an attempt was made to include this information for all participants, even beyond the last wave of testing.

The present study includes all individuals for whom more than 50% of demographic information was available and who completed more than 50% of the cognitive and health items for at least one wave of the study including at least the first MMSE. Forty-eight participants were excluded because they did not complete a sufficient portion of the study protocol at any wave and two were excluded because no MMSE was completed at the first assessment wave. A single outlier was excluded based on their observed trajectory showing a large increase in MMSE score from the first to second occasion. Failure to complete the MMSE and sufficient portion of the study protocol appears to be due to compromised health, either because of dementia or general frailty. Individuals not included in the subsample analyzed here were more likely to be diagnosed with dementia at, or prior to, the first assessment. The aim of the present study is to determine whether MMSE scores in a given individual

over time are more telling of dementia status than cut off methods based on a single MMSE result.

Although the majority of individuals excluded from the sample by these criteria were those who had been diagnosed with dementia, we felt that, given the inability of these individuals to complete the majority of the assessment, they are unlikely to be representative of participants in other studies.

Individuals whose health has declined to the extent that they are unable to complete the MMSE or most other assessment measures are also unlikely to benefit from efforts to improve the utility of the MMSE as a screening measure for dementia as they were likely diagnosed some years before.

Procedure

Participants were assessed by registered nurses trained for the study and regularly supervised. Testing sessions were conducted in the participant's residence and normally lasted 3.5 to 4.0 hours including several rest periods. Twin pairs were assessed by different nurses and within 1 month of the co-twin's test session. Scheduling was done to minimize geographical, age, and gender order effects across participants. Nurses were blind to the zygosity of the twins to avoid expectation bias. When the zygosity of the twin pair was not already known, it was determined by DNA test.

Measures

The OCTO Twin Study included a large battery of measures including assessment of health and functional capacity, personality, well-being, and interpersonal functioning. For the present analysis only a subsample of these measures is used.

Demographic Information

Demographic information was collected from participants during the first assessment. When possible, name, year of birth, and in some cases level of education completed were verified through the information contained in the Swedish Twin Registry. Date of death was obtained through the Death Registry.

Mini Mental State Exam (MMSE; Folstein et al.,1975).

A Swedish version of the MMSE was administered. The version includes the Swedish words for key, toothbrush, and lamp in place of apple, penny, table for immediate and delayed recall. The English translation of the phrase used instead of “No ifs ands or butts” is “burned down two-family house”. In this study, Serial 7’s was used without an option for spelling a word backwards. Several other changes from the English version were that the instruction was “point at the door” instead of “close your eyes” and participants were instructed to put the folded paper on their laps instead of the floor. Examiners were instructed to record whether a question was not answered due to sensory or motor impairments.

Cognitive Functioning.

Cognitive functioning was assessed with a Swedish psychometric battery which is a commonly used set of tests intended to assess memory and abilities from both the fluid and crystallized domains (Dureman & Sälde, 1959).

Dementia.

Individuals suspected of having dementia were diagnosed through discussions of a multidisciplinary team in which all available information was used in the diagnostic process, including a review of medical records, in-person testing protocols, and an informant interview about memory and cognitive problems (Pedersen, Gatz, Berg, & Johansson, 2004). Dementia was diagnosed according to criteria of the DSM-III-R (American Psychiatric Association, 1987) and includes dementia due to Alzheimer disease and vascular dementia, mixed and secondary dementia, as well as unclassified cases meeting the criteria.

Data Analysis**Sensitivity and specificity**

Sensitivity is a commonly used method for evaluating the ability of a diagnostic or screening tool to identify those who have the target condition. The number of people identified by the new

method is compared to the number known to have the condition (“known” using another more established or “gold standard” method). Those correctly identified as having the condition are referred to as true positives. Those incorrectly categorized as not having the condition when they do are false negatives. Using the traditional cut-off score of less than 24, the sensitivity of the MMSE was calculated for the first assessment wave of the current sample. Sensitivity, calculated as the number of individuals diagnosed with dementia by the first assessment who were correctly identified by the traditional MMSE cut-off (true positives), divided by the total number of participants diagnosed with dementia by the first assessment. Sensitivity using the cut-off score method provides a baseline against which the accuracy of the group-based trajectory models in detecting individuals with dementia can be compared. A second sensitivity calculation was conducted to determine the sensitivity of the MMSE cut-off in identifying individuals who were diagnosed by or at the fifth assessment wave. The same calculation was used: the number of individuals correctly identified by the MMSE cut-off (<24) at the fifth wave of assessment, divided by the number of participants diagnosed with dementia by or at the fifth wave of assessment. Sensitivity was calculated using MMSE scores at the fifth wave because that information is included in the group-based trajectory models as well.

Specificity calculations identify how many people are correctly classified as not having a condition of interest by comparing the number of people not identified, using the measure in question, with the number of people identified as not having the target condition as determined by the accepted standard. Those who are correctly identified as not having the condition of interest are referred to as true negatives. Individuals who have not been diagnosed with the condition but are categorized as having it by the screening measure are referred to as false positives. Using the traditional MMSE cut-off score of less than 24, specificity of the measure in our sample was calculated. Specificity is the number of participants not classified by the MMSE cut-off score (i.e., those with a MMSE score of 24 or greater) as having dementia (true negatives) divided by the total number of participants without

dementia in the sample. Again, the same calculation was conducted using MMSE scores and diagnostic status at baseline and then again with fifth assessment wave MMSE scores and diagnostic status. Specificity is a measure of how well individuals without dementia are identified. Thus, calculating it using cut-off scores provides a baseline against which to compare how well the group-based trajectory models identify such individuals.

Positive predictive value and negative predictive value

Positive predictive value is the proportion of positive results that are true positives. This is calculated as the number of individuals who scored less than or equal to 24 and have been diagnosed with dementia by or at the first assessment, divided by the total number of individuals who scored less than 24 on the MMSE. As with specificity, the veracity of results is based on a comparison to a more established “gold standard”. Positive predictive value provides the percentage of individuals identified by the screening measure as having dementia, that actually have dementia. This provides a measure of how confident you can be that a person has dementia, if you know they scored less than or equal to 24 on the MMSE. Comparing the values obtained using cut-off scores and those obtained using the group-based trajectory model provides another way to ascertain whether one is superior.

Negative predictive value is the proportion of individuals whose test results indicate they are free of the condition of interest who are actually free of the condition. Negative predictive value is calculated as the number of individuals whose MMSE score is greater than 24 who are not diagnosed with dementia by or at the first assessment wave (true negatives) divided by the total number of individuals with MMSE greater than 24. This is similarly calculated for the first and fifth waves of assessment as a point of comparison to evaluate the success of group-based trajectory models.

Logistic regression

Initially, a two-level logistic regression analysis using data from the first wave of the study was conducted to determine whether the prediction of dementia could be improved by the inclusion of

predictor variables in addition to the MMSE. A two-level logistic regression analysis controls for the lack of independence of twin pairs in the analysis. However, to compare whether twin pair clustering affected the outcome, a single level logistic regression was also conducted. Chronological age, years of formal education, and sex were added to MMSE score as predictors of the categorical outcome diagnosed with dementia by time 1 or not in both logistic regression analyses. The results of the two-level logistic regression were not different from a standard logistic regression. Due to the increased complexity of interpreting results from a multilevel logistic regression, only the results of the standard single level logistic regression will be presented.

Trajectory modeling

The sample size required for a growth mixture analysis is typically quite large, but in a Monte Carlo simulation study a sample size of 300 had acceptable results, though with reduced power relative to a sample size of 3000 (Muthén, 2004). Therefore, it should be feasible to conduct a growth mixture model analysis with the current sample.

In the present analysis, we began with the simplest possibility as the null hypothesis: that a single unconditional latent growth curve characterizes the MMSE scores of participants over time. Then a series of growth mixture models, with MMSE scores as the continuous dependent variable, were compared first to the single class model and then to one another. Developing a “final” model is an iterative process. Generally, the process consists of two stages: the first, examining unconditional growth mixture models to determine the optimal number of trajectory classes, and a second stage, where covariates are included and the model including covariates is compared to the unconditional model. For the first stage, we used an adaptation of the common approach of comparing the fit indices of models with successive numbers of classes to arrive at the model with the best fit (Jung & Wickrama, 2008; Muthén, 2004; Petras & Masyn, 2010).

First, a simple unconditional latent growth curve was fit to examine changes in MMSE scores over time and look for possible distributions and the shape of change trajectories. The unconditional growth model was in the form:

$$\text{MMSE}_{ij} = \beta_{0j} + \beta_{1j} \text{TIME}_{ij} + \beta_{2j} \text{TIME}_{ij}^2 + r_{ij}$$

Where MMSE_{ij} is the MMSE score of person j at time i . β_{0j} is the intercept, the MMSE score of individual j at Time 0, in this case specified as the start of the study. β_{1j} is the slope for person j , β_{2j} is the quadratic term for person j . r_{ij} is a person specific residual. Another unconditional latent growth curve, the same except for it included only the linear growth term was also explored.

Next, the two-class model was specified and fit indices were compared to the one-class model. Models were evaluated in four ways, by the Bayesian Information Criterion (BIC; Schwarz, 1978), the Vuong-Lo-Mendell-Rubin test (VLMR-LRT; Lo, Mendell, & Rubin, 2001), the Bootstrapped Likelihood Ratio Test (BLRT), and model entropy. For each model the BIC is calculated as:

$$\text{BIC} = \log(L) - 0.5 * \log(n) * (k)$$

L is the value of the model's maximized likelihood, n is the number of participants, and k is the number of parameters in the model. A larger value indicates better model fit, and because the BIC is always negative this means selecting the least negative value, often the minus sign is simply omitted in which case smaller numbers are better. For a model that fit the data perfectly, the BIC would be 0. The term $\log(L)$ is always negative, and is reduced when there are more parameters in the model. The second half of the equation subtracts a penalty for the number of parameters so that adding a trajectory group (additional parameters) is only good if the improvement in model fit is greater than the penalty for more parameters (the second term). The BIC can be used to compare both nested and not nested models and so is particularly helpful for our purposes (Nagin, 1999). Nagin (1999) found that in comparing models with different numbers of classes the BIC reaches a "maximum". That is, the BIC improves with increasing number of classes but reaches a point at which models with additional classes

have lower BIC values. However, it has been noted that arriving at a clear highest value BIC before it begins to decrease again with each additional class does not always happen before including so many classes that the model fails to converge (Petras & Maysn, 2010). In some cases it is necessary to look for diminishing returns rather than a clear indication that the k-class model is superior (Petras & Maysn, 2010).

The VLMR-LRT is a likelihood ratio test that can be used to compare nested models. Log likelihood difference tests compare model fits using a chi-square difference distribution to test the significance of the difference (Nylund, Asparoutiov, & Muthen, 2007). However, likelihood ratio tests cannot be used to compare the likelihood of a k-1 class and a k class model, where k is the number of classes, because the difference does not follow a chi-square distribution. The VLMR-LRT, implemented in MplusV5.1 software (Muthén & Muthén 1998-2008b), compares a k-class model to a (k-1)-class model (the null model) and provides a *p*-value to indicate whether the k class model is a significant improvement.

The BLRT is similar to the VLMR-LRT in that both are likelihood ratio tests that provide a *p*-value indicating whether a k-class model is superior to a k-1 class model. However, the BLRT does not assume that the difference follows a known distribution. Instead the BLRT uses bootstrap samples to empirically estimate the distribution of the log likelihood difference test statistic (Nylund, et al., 2007). In their Monte Carlo simulation study of latent class enumeration in a growth mixture modelling, Nylund et al. (2007) found the BLRT had the best odds of identifying the correct number of classes.

Posterior probabilities are calculated for each individual and indicate the probability of said individual falling into each class in the model. Posterior probabilities range from 0 to 1 with higher values indicating greater certainty of the individual being placed in one class versus the others. Entropy is a summary measure calculated from the posterior probabilities of the whole sample that indicates the degree to which the latent classes are distinguished from one another and the certainty with which

individuals are placed in classes. Higher entropy indicates better separation of classes. Although not a measure of model fit, entropy does indicate the degree to which each class represents a homogenous trajectory group that is different from the others (Petras & Maysn, 2010).

As described previously, the difference between latent class growth models and growth mixture models is that for the former, allowing different growth parameters for each class is intended to capture variance within the sample, and thus variance around the growth parameters within each class is constrained to be zero. In growth mixture models, variance around growth parameters can be modeled, it can be specified to be the same across all classes or allowed to be different for each class. Residual variance can similarly be constrained to be equal or allowed to differ across classes. The choice of how to specify the model is influenced by several factors. A first, practical consideration is that allowing residual variances to differ across classes increases the chances of nonconvergence because the likelihood function is not bound and the resulting model may include variances of zero and classes with only one individual (Petras & Masyn, 2010). Thus, the current analysis begins by constraining residual variance to be equal across classes, but allowing variance around growth parameters (intercept, linear slope, and quadratic slope) to differ. The shape of the change trajectories was determined by first examining an unconditional model with k classes and only an intercept and linear slope term and then an identical model that also included a quadratic term.

Following Petras and Masyn (2010), the second stage of model building examined the effects of the covariates age, years of education, and sex. The addition of covariates to growth mixture model specification can alter the number and formation of estimated classes (Huang, Brecht, Hara, & Hser, 2010). The effects of the covariates were allowed to predict class membership first. Petras and Masyn (2010) suggest that if the addition of covariates results in substantial differences in either the proportion of individuals in each class or in the growth parameters this may signal a misspecification of covariate relations with latent class indicators. They suggest that only in that case should covariate effects on the

growth parameters themselves be modeled. For comparative purposes, models with the effect of covariates on class membership, and on the growth parameters (intercept, linear slope, and quadratic slope) of each class, were also examined. The same procedure of sequentially comparing models starting with a single class model and adding a class until a four class model was run or the differences in model fit indices were not significant was followed.

In group-based trajectory modeling, determining the best model is a matter of making relative comparisons to other models, there are no absolute tests of model fit. However, in the present study, predicting whether the individual is or will be diagnosed with dementia or not is the question of interest and provides an excellent further test of the “best” fitting models. As described above, the posterior probability is calculated for each individual, who is then grouped into the class for which they have the highest posterior probability. The certainty of membership in each class was examined as an indication of model fit. Sensitivity and specificity of class membership is examined using the available dementia diagnosis information. Further, the wave of assessment in which diagnosis of dementia occurred is explored in comparison to class membership.

Results

Participants were on average 83.40 years old at the first assessment, the majority were female (65%), with an average of 7 years of education, and nearly 10% were diagnosed with dementia by or at the first assessment. Demographic information and frequency of dementia at first assessment and over the study period is presented in Table 1.

Table 1. Demographic information and dementia status of participants.

		Participants (<i>n</i> = 651)
		M (SD)
Age at first interview		83.40 (3.10)
Years of education		7.14 (2.30)
		<i>n</i> (%)
Sex		
	Female	425 (65%)
	Male	226 (35%)
Diagnosis of dementia at assessment 1		
	None	591 (91%)
	Probable	60 (9%)
Diagnosis of dementia over the study period		
	None	465 (71%)
	Probable	186 (29%)

Sensitivity and specificity

Using the traditional MMSE cut-off of 24 to detect dementia at the first assessment resulted in a sensitivity of 80%, specificity of 85%, positive predictive value of 36%, and negative predictive value of 98% (Table 2). Using the same cut-off score to detect dementia at the last wave of assessment resulted sensitivity of 97%, specificity of 69%, positive predictive value of 39%, and a negative predictive value of 99% (Table 2).

Table 2. Cross-tabulation of classification based on formal diagnosis and on MMSE less than 24 proxy at first and fifth assessments.

MMSE Score	Time 1		Time 5	
	Cases of Dementia	Cognitively Intact	Cases of Dementia	Cognitively Intact
Less than 24	48	87	36	56
24 or Greater	12	504	1	126

Logistic regression

A full model logistic regression with all three predictors against a constant-only model was statistically significant, $\chi^2(4, N = 651) = 163.41, p < .001$. This indicates that the set of predictors reliably distinguished between individuals with dementia and those without. However, according to the Wald criterion, only MMSE reliably predicted dementia status, $\chi^2(1, N = 651) = 85.1, p < .001$. Years of education and age did not help in predicting dementia status, $\chi^2(1, N = 651) = 2.01, p = .157$, and $\chi^2(1, N = 651) = .02, p = .888$, respectively.

The model correctly classified 94% of cases overall, using the standard 0.5 probability cut-off. Specificity, the classification of individuals without dementia, was 99%, with only seven individuals without dementia misclassified as having dementia. However, the MMSE-based classification of individuals with dementia was less reliable, with a sensitivity of 45%. Of those with dementia, 27 were correctly classified (true positives), while 33 were incorrectly predicted to not have dementia (false negatives). The positive predictive value was high at 79% although not as high as the negative predictive value, which was 95%.

Unconditional growth mixture models

The analysis of trajectories of change in MMSE scores began with an unconditional one-class model. The one-class model is a simple linear random effects model of MMSE scores over time. A one-class model with only linear change was modeled as well as a one-class model with both linear and

quadratic slope terms. The significant quadratic slope term, and an examination of a selection of individual trajectories (Figure 1) indicate that a quadratic trajectory describes the shape of change in MMSE scores for at least portion of individuals. The lower BIC value also suggests that the unconditional latent growth curve with a quadratic shape is a better fit. Therefore linear and quadratic change parameters were included in all subsequent models. The fit indices and parameters for the unconditional models including linear and quadratic terms are shown in Table 3 and Table 4, respectively.

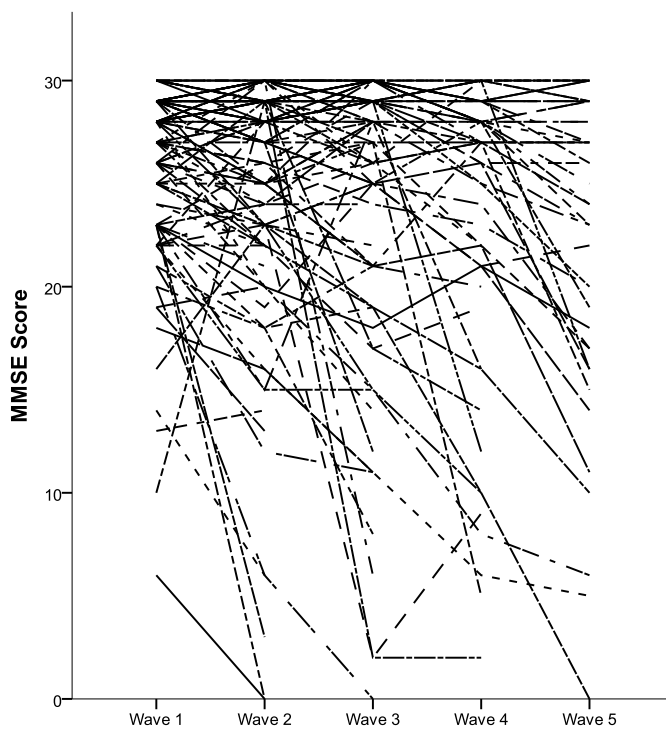


Figure 1. A random selection of individual observed trajectories.

As is common in growth mixture models, residual variances were constrained to be equal across classes (Petras & Masyn, 2010). Relaxing this constraint confirmed that convergence problems arise without it. The residual variance was allowed to vary for the MMSE scores at each measurement occasion. However, in the two and three class models this resulted in improper convergence due to a negative residual variance. Following suggested procedures for addressing such issues, the negative residual variance was constrained to zero in both the two and three class conditional models. All three

fit indices indicate a significantly better fit for the two-class model than the one-class model: a smaller BIC, and a significant VLMR-LRT and BLRT (see Table 3). The entropy value indicates that separation between the two classes is excellent. Most participants were assigned to class one ($n = 608$), which had a higher intercept than class one and significant linear (Est = -0.62) and quadratic (Est = -0.06) slopes. Class two in contrast, was much smaller ($n = 43$), had an intercept 15 points lower, and had non-significant linear and quadratic slope terms (see Table 4). Average scores at each occasion for the two classes are displayed in Figure 2.

Table 3. Fit indices for unconditional models.

Model	# of free parameters	BIC	VLMR-LRT	BLRT	Entropy	Smallest class (n)
1-class	14	12576.796	n/a	n/a	n/a	n/a
2-class	18	12290.843	$p = .0244$	$p < .001$	0.97	43
3-class	21	12046.893	$p < .001$	$p = 1.0$	0.90	35

Table 4. Parameter estimates of the unconditional quadratic models.

	One-class model		Two-class model				Three-class model					
	N = 651		High Performing Slow Decline N = 608		Stable Low Performance N = 43		High Performing Late Decline N = 522		Rapidly Declining N = 94		Decreasing Low Performance N = 35	
Fixed Effects	b	SE	b	SE	b	SE	b	SE	b	SE	b	SE
Intercept	25.80**	(0.19)	26.80**	(0.19)	11.31**	(1.84)	27.30**	(0.15)	23.79**	(0.57)	10.41**	(1.27)
Time	-0.68**	(0.11)	-0.62**	(0.14)	-1.42	(1.24)	0.27**	(0.07)	-4.77**	(0.49)	-1.97**	(0.60)
Time ²	-0.06**	(0.01)	-0.06*	(0.02)	-0.02	(0.24)	-0.12**	(0.01)	0.21*	(0.07)	0.11	(0.09)
Variance Components	b	SE	b	SE			b	SE				
Intercept	23.05 **	(2.14)	8.78**	(1.60)			7.49**	(1.52)				
Time	4.11 **	(0.57)	4.09**	(0.80)			0.78*	(0.36)				
Time ²	0.04**	(0.01)	0.04*	(0.01)			0.02**	(0.01)				
Cov (ILS)	2.77 *	(0.99)	1.81*	(0.84)			-0.58	(0.69)				
Cov (IQS)	-0.10	(0.11)	-0.08	(0.11)			0.11	(0.07)				
Cov (LSQS)	-0.31**	(0.07)	-0.31**	(0.09)			-0.10*	(0.04)				
Residual Variances												
MMSE 1	1.07	(1.61)	0.98	(1.50)			1.09	(1.46)				
MMSE 2	8.90**	(0.98)	8.88**	(1.55)			10.12**	(1.52)				
MMSE 3	7.17**	(1.11)	7.15**	(1.81)			5.26**	(1.10)				
MMSE 4	6.23**	(1.13)	6.18**	(1.80)			6.92**	(1.35)				
MMSE 5	1.14	(2.96)	1.10	(5.15)			0	(0)				

Note: ILS = Intercept*Linear slope, IQS = Intercept*Quadratic slope, LSQS = Linear Slope*Quadratic slope
* $p < .05$; ** $p \leq .001$

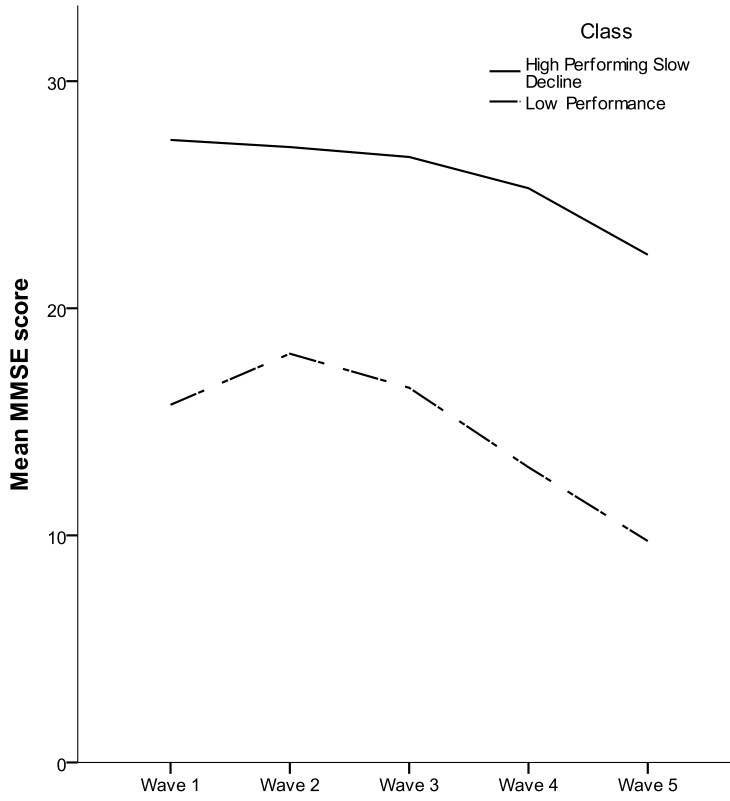


Figure 2. Average scores for the classes in the two-class unconditional quadratic model.

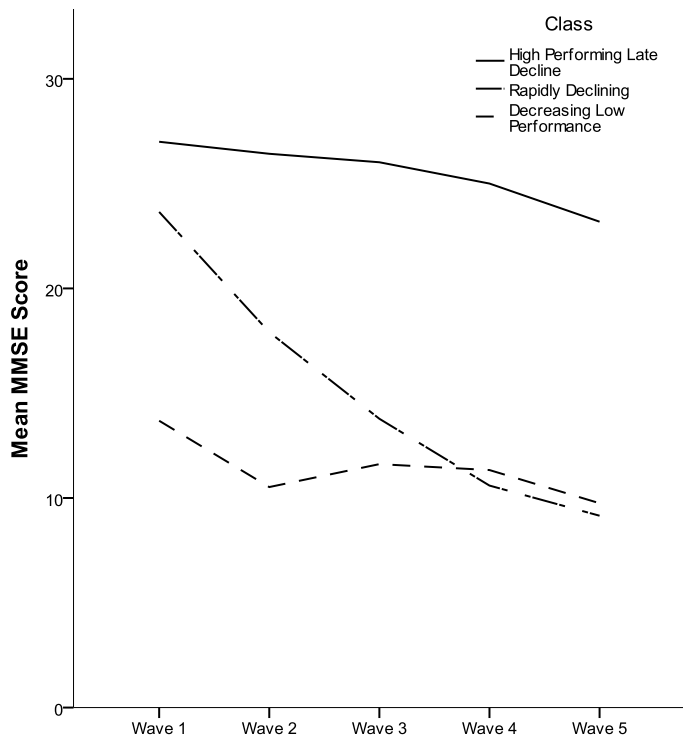


Figure 3. Average scores for the classes in the three-class unconditional quadratic model.

Parameter estimates of the three-class unconditional model including quadratic change are shown in Table 4, and fit statistics in Table 3. The three-class model provided a better statistical fit to the data according to the BIC (12046.893 vs. 12290.843) and the VLMR-LRT however, according to BLRT, it did not. Mean MMSE scores for each of the classes at each occasions in the three-class unconditional quadratic model are presented in Figure 3. A four-class unconditional model was run but the conditional four class model would not converge properly due to significant negative residual variances so neither four class model was considered. This can be taken as suggesting that a model with a smaller number of classes is more appropriate (Nylund et al. 2007).

Conditional models

The effects of the covariates, sex, age, and years of education, were allowed on class membership. Beginning with a one-class model, as for the unconditional models, a series of models was tested. As in the unconditional model series, both linear and quadratic change terms were included in the model specification.

Parameter estimates for the conditional one-class quadratic model are presented in Table 5. These are presented separately from those of the other conditional models because covariate effects were modeled on growth parameters directly; in subsequent models (with more than one class) covariate effects were only modeled on class membership. The results of the one-class model indicate a statistically significant linear slope parameter and a non-significant quadratic slope. However, significant variance around the quadratic term suggests that it should be retained. There were no significant effects of sex on intercept, linear slope, or quadratic slope terms. There was a significant effect of age on initial MMSE scores and linear slope, but not quadratic slope. The effect of years of education was significant on the intercept but not the linear or quadratic slope terms.

Table 5. Parameter estimates of the one-class conditional quadratic model.

One-class model		
N = 651		
Fixed Effects	b	SE
Intercept	23.10**	(0.69)
Time	-1.12*	(0.40)
Time ²	-0.09	(0.05)
Sex	0.32	(0.40)
Age	-0.25**	(0.06)
Years of Education	0.36**	(0.08)
Time*Sex	-0.02	(0.23)
Time*Age	-0.08*	(0.04)
Time*Years of Education	0.07	(0.05)
Time ² *Sex	0.04	(0.03)
Time ² *Age	0.00	(0.01)
Time ² *Years of Education	0.00	(0.01)
Variance Components		
Intercept	21.60**	(2.10)
Time	4.02**	(0.56)
Time ²	0.04**	(0.01)
Cov (ILS)	2.53*	(0.98)
Cov (IQS)	-0.13	(0.11)
Cov (LSQS)	-0.31**	(0.07)
Residual Variance		
MMSE 1	1.14	(1.62)
MMSE 2	8.98**	(0.99)
MMSE 3	7.26**	(1.10)
MMSE 4	5.91**	(1.10)
MMSE 5	1.64	(2.92)

Note: ILS = Intercept*Linear slope

IQS = Intercept*Quadratic slope

LSQS = Linear Slope*Quadratic slope

* $p < .05$; ** $p \leq .001$

Table 6. Fit indices for conditional quadratic models.

Model	# of free parameters	BIC	VLMR-LRT	BLRT	Entropy	Smallest class (n)
1-class	23	19680.427	n/a	n/a	n/a	n/a
2-class	20	12296.724	$p = .0155$	$p < .0001$	0.968	42
3-class	27	12057.561	$p = .025$	$p = .666$	0.899	34

As in the unconditional model series, the residual variance for MMSE scores were allowed to vary across occasion but not class except for the fifth assessment MMSE score which was constrained to be 0. A negative residual variance resulted when the constraint was not applied. The two-class quadratic model allowing the covariate effects on class showed a significantly better fit than the one

class model by all three fit indices: the BIC was lower (12296.724 vs. 19680.427) and the VLMR-LRT and BLRT were significant (see Table 6). The entropy value for the two-class model indicated very good separation of the two classes (entropy = 0.968). Parameter estimates for the two-class and three-class conditional models are presented in Table 7. In the two-class model the majority of participants were assigned to class one (n = 609). Class one trajectories are described by a high mean intercept (Est. = 26.80), a significant linear slope (Est. = -0.62), and a significant quadratic slope (Est. = -0.06). Class two averaged a much lower intercept (Est. = 11.35) and non-significant linear and quadratic slope terms. The covariates sex and age were not significantly related to class membership (both p 's > .05). Years of education was significantly related to class membership such that individuals assigned to class one were likely to have greater years of education.

Table 7. Parameter estimates for the two and three-class conditional quadratic models.

	Two-class model				Three-class model					
	High Performing Slow Decline		Low Performance		High Performing Late Decline		Rapidly Declining		Decreasing Low Performance	
	N = 609		N = 42		N = 513		N = 104		N = 34	
Fixed Effects	b	SE	b	SE	b	SE	b	SE	b	SE
Intercept	26.80**	(0.19)	11.35**	(1.93)	27.37**	(0.24)	23.67**	(0.69)	10.32**	(1.31)
Time	-0.62**	(0.13)	-1.39	(1.03)	0.27**	(0.07)	-4.61**	(0.89)	-1.97*	(0.65)
Time ²	-0.06**	(0.02)	-0.03	(0.19)	-0.12**	(0.01)	0.19	(0.10)	0.11	(0.10)
Variance Components	b	SE			b	SE				
Intercept	8.90**	(1.66)			7.36**	(1.63)				
Time	4.15**	(0.76)			0.84	(0.51)				
Time ²	0.04**	(0.01)			0.02**	(0.01)				
Cov (ILS)	1.80*	(0.79)			-0.72	(0.88)				
Cov (IQS)	-0.08	(0.09)			0.12	(0.07)				
Cov (LSQS)	-0.32**	(0.08)			-0.10	(0.06)				
Residual Variances	b	SE			b	SE				
MMSE 1	0.89	(1.55)			0.99	(1.48)				
MMSE 2	8.89**	(1.54)			10.26**	(1.58)				
MMSE 3	7.01**	(1.77)			5.33**	(1.15)				
MMSE 4	6.44**	(1.23)			6.91**	(1.37)				
MMSE 5	0	(0)			0	(0)				

Note: ILS = Intercept*Linear slope, IQS = Intercept*Quadratic slope, LSQS = Linear Slope*Quadratic slope
* $p < .05$; ** $p \leq .001$

The fit of this three-class model exceeded that of the two-class model according to two of the three fit indices: the BIC was lower, and the VLMR-LRT was significant ($p = .025$), the BLRT was not

significant ($p=.666$). However, entropy was greater for the smaller model (0.968 vs. 0.899). The four-class model was not examined due to model convergence issues that were not resolved by setting negative variances to zero. In the three-class model, there was a class with a high mean initial MMSE (Est. = 27.37), a significant linear slope term, indicating a significant instantaneous rate of change at intercept (Est. = 0.27), and a significant quadratic term indicating negative curvature (Est. = -0.12), tentatively labeled the High Performing Late Decline class. The second class had a lower intercept (Est. = 23.67), negative instantaneous rate of change (Est. -4.61), and significant positive curvature (Est. = 0.19) and so was labeled the Rapidly Declining class. The average trajectory of the third class, characterized by low initial mean MMSE scores (Est. = 10.32), a significant negative linear slope at the intercept (Est. = -1.97), and non-significant curvature (Est. = 0.11), suggested the label Decreasing Low Performance. Average scores for each class in the three-class conditional quadratic model over the five waves can be seen in Figure 5. Observed trajectories for all individuals in each of the three classes are displayed in Figures 6, 7, and 8.

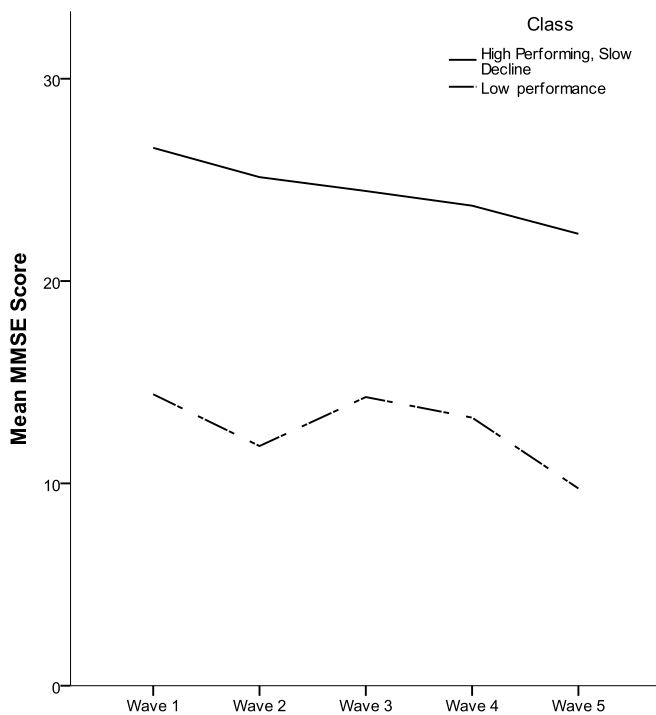


Figure 4. Average scores for the classes of the two-class conditional quadratic model.

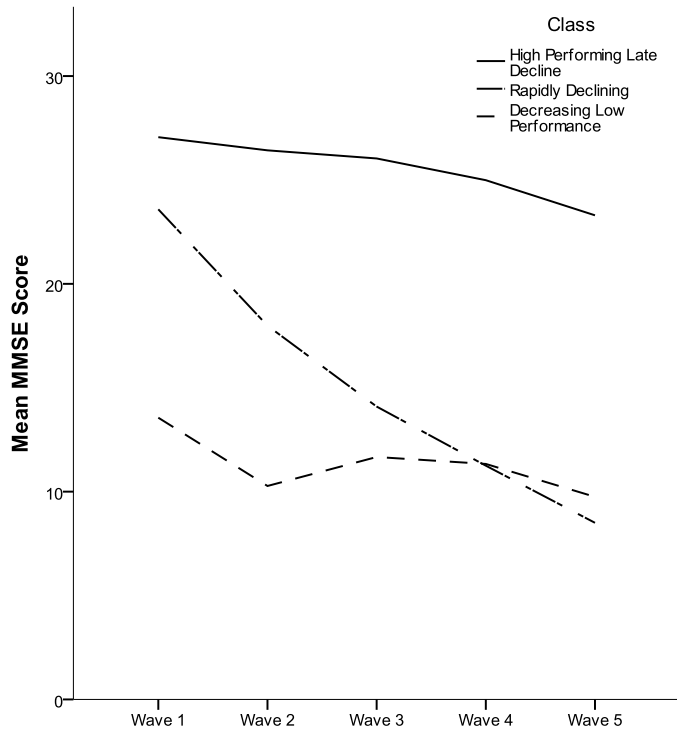


Figure 5. Average scores for each of the classes in the three-class conditional quadratic model.

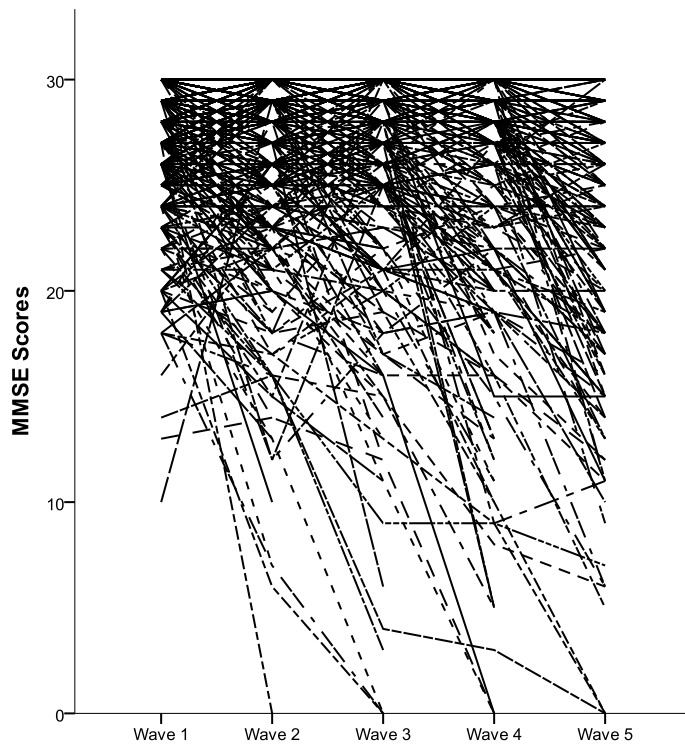


Figure 6. Observed trajectories of individuals in the "High performing Late Decline" class of the three-class conditional quadratic model.

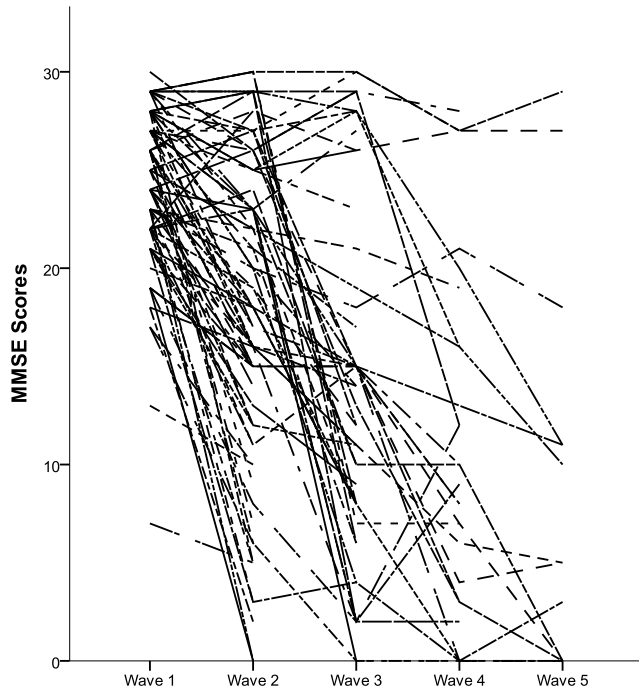


Figure 7. Observed scores of individuals in the "Rapid Decline" class of the three-class conditional quadratic model.

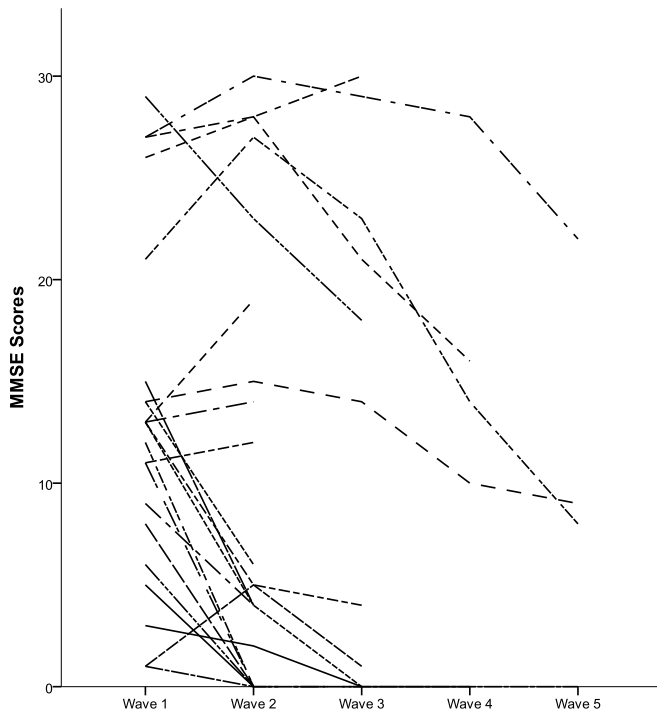


Figure 8. Observed scores of individuals in the "Decreasing Low Performance" class of the three-class conditional quadratic model.

In the three-class model, sex did not significantly distinguish between any of the classes (See Table 8). Years of education significantly distinguished the “High Performing Late Decline” class from the “Decreasing Low Performance” class, with individuals with higher levels of education more likely to be in the “High Performing Late Decline” class. Age only distinguished the Rapidly Declining class from the High Performing Late Decline class such that older individuals were more likely to be in the Rapidly Declining Class.

Table 8. Covariate effects on the three-class quadratic model.

Covariate	Target class	Reference class	Est.	S.E	<i>p</i> - value	Est. OR ^a
Sex	Rapidly Declining	Decreasing Low	-0.120	0.456	0.792	0.89 ^b
	High Performing	Decreasing Low	0.015	0.394	0.970	1.02 ^b
	Rapidly Declining	High Performing	-0.135	0.263	0.608	0.87 ^b
Years of Education	Rapidly Declining	Decreasing Low	0.160	0.104	0.125	1.17 ^c
	High Performing	Decreasing Low	0.295	0.087	0.001*	1.34 ^c
	Rapidly Declining	High Performing	-0.135	0.087	0.120	0.87 ^c
Age	Rapidly Declining	Decreasing Low	0.082	0.066	0.214	1.09 ^d
	High Performing	Decreasing Low	-0.069	0.060	0.250	0.93 ^d
	Rapidly Declining	High Performing	0.151	0.041	< .0001*	1.16 ^d

High Performing = High Performing Late Decline, Decreasing Low = Decreasing Low Performance

^a Odds (membership in target class): Odds (membership in reference class) among individuals in either target or reference class

^b Calculated for Sex = Female vs. Sex = Male

^c Calculated for a 1 SD increase in education

^d Calculated for a 1 SD increase in Age

* Statistically significant

Comparing across these two series, the unconditional and conditional quadratic models are very similar in both the proportion of the sample assigned to each class and the meaning of the classes. The three-class model was significantly better than the two-class model for both unconditional and conditional model series by two out of the three fit indices. Looking at the conditional models, while the variance of the intercept in the two-class model (8.90) is smaller than in the single class (21.60), the variance terms of the linear (4.15) and quadratic (0.04) slope terms for the two-class models are not much smaller than those in the single class models (linear slope variance = 4.02, quadratic slope

variance = 0.04) indicating that when only two classes are allowed the classification is based primarily on intercept rather than slope. The smaller variance of the slope terms (linear slope variance = 0.84, quadratic slope variance = 0.02), particularly the linear slope, of the three-class model suggests that with three classes, different trajectories are better grouped into separate classes. This holds true for the unconditional models as well, which, along with the fact that the Rapidly Declining class does seem to represent a group of participants with substantially different trajectories, suggests that the three-class models are more appropriate.

Classification Accuracy

The meaning of the latent classes formed in the models thus far can be inferred from the mean initial MMSE scores and the shape of the decline in MMSE scores over time. However, investigating whether membership in a particular class corresponds to whether that individual has, or will develop dementia is an essential component of testing our hypothesis. The present analysis is unique in having such criteria to aid in the assessment of how well the latent trajectory models describe the data. The correspondence between class membership and actual diagnosis of dementia was examined for four of the models: the two-class unconditional quadratic model, the three-class unconditional quadratic model, the two-class conditional quadratic model, and the three-class conditional quadratic model. Although, the three-class models appeared superior, we also examined the predictive accuracy of the two-class model, because previous studies have found reasonable prediction of dementia status from a two-class model (Small & Bäckman, 2007), and we did not want to prematurely rule it out.

The classification accuracy of two-class models can be easily described in terms of sensitivity, specificity, positive predictive value, and negative predictive value. Class one is viewed as representing individuals who remain dementia free over the study period, analogous to that in the three-class models. The second class is viewed as representing those who have or will develop dementia for both the conditional and unconditional models. Model based classifications compared to diagnostic

outcomes for the two-class unconditional quadratic model are presented in Table 9. Only 15 percent of individuals who had or developed dementia were correctly assigned to the Low Performance class, meaning the sensitivity of the model is 15%. Individuals who would remain dementia free were correctly assigned 97% of the time to the High Performing Slow Decline class, which appears to represent individuals without dementia, thus the model specificity is 97%. Positive predictive value is 65% and negative predictive value is 74%. This indicates that if an individual is assigned to the Low Performance class they have a 65% chance of actually having dementia and if an individual has been assigned to the High Performing Slow Decline class there is a 74% chance that person has remained free of dementia.

The classification accuracy of the two-class conditional quadratic model in detecting individuals who had or would develop dementia was the same in terms of sensitivity, specificity, and negative predictive value, and was only one percent lower in positive predictive value (64% vs.65%). Table 9 displays a cross-tabulation of dementia status predicted by class membership (High Performing Slow Decline, Low Performance) compared to “actual” dementia status.

Table 9. Cross-tabulation of unconditional and conditional two-class model classification and dementia diagnostic status.

	Unconditional model		Conditional model	
	Cases of Dementia	Cognitively Intact	Cases of Dementia	Cognitively Intact
High Performing Slow Decline	158	450	159	450
Low Performance	28	15	27	15

Given the poor sensitivity of the two-class models and the apparently better fit of the three-class models, the latter were examined in more detail. Three class models are less amenable to descriptions

in terms of sensitivity and specificity. For the unconditional three-class model, considering the Rapid Decline class and the Declining Low Performance class together as individuals identified as having dementia by the model, and comparing that to actual diagnosis over the course of the study, the sensitivity is 46%, specificity is 91%, positive predictive value is 66%, and negative predictive value is 81%. In comparing the three-class models, correct classification of individuals diagnosed with dementia was very similar between the unconditional and conditional models except that a slightly greater percentage of individuals with dementia were assigned to the Rapidly Declining group (63% vs 57%) in the unconditional model (see Table 10 and Table 11). This is because there are fewer individuals without dementia in the Rapidly Declining class of the unconditional model but the number of individuals with dementia is the same, suggesting better classification.

At each wave of the study, the number of individuals diagnosed with dementia, and the number who remained in study (with and without a diagnosis of dementia) was examined for each of the three-class models. The number of individuals who have remained in the study at each wave and the percentage of the initial membership of each class at baseline who are still in study is given in Table 10. Table 10 also shows the number of individuals who have a diagnosis of dementia at each wave with those diagnosed as a percentage of the “surviving” participants in brackets. Table 11 shows the same for the three-class conditional quadratic model. Classes differed in the proportion of individuals still in the study at each wave. Individuals in the High Performance Late Decline class were more likely to remain in the study through to the fifth wave than those in the other two classes (40% in both the unconditional and conditional models). The percentage of individuals in the Rapidly Declining class who were still in the study at wave 5 was much lower (14% in unconditional, 13% in conditional model), and the percentage of individuals in the Decreasing Low Performance class were even less likely to still be in the study at wave 5 (11% in the unconditional, 12% in the conditional model).

Table 10. Three-class unconditional quadratic model diagnosis of dementia and survival in study by wave.

	High Performing Late Decline (Time 1: n = 522)		Rapidly Declining (Time 1: n = 94)		Decreasing Low Performance (Time 1: n = 35)	
	Diagnosed	Survived	Diagnosed	Survived	Diagnosed	Survived
	(% of survived)		(% of survived)		(% of survived)	
Time 1	17 (3%)	-	19 (20%)	-	24 (69%)	-
Time 2	29 (6%)	463 (89%)	41 (49%)	83 (88%)	19 (73%)	26 (74%)
Time3	36 (10%)	355 (68%)	34 (68%)	50 (53%)	8 (62%)	13 (37%)
Time 4	38(14%)	274 (52%)	15 (56%)	27 (29%)	4 (67%)	6 (17%)
Time 5	27 (13%)	207 (40%)	7 (54%)	13 (14%)	3 (75%)	4 (11%)
Ever	101 (19%)	40%	59 (63%)	14%	26 (74%)	11%

Table 11. Three-class conditional quadratic model diagnosis of dementia and survival in study by wave.

	High Performing Late Decline (Time 1: n = 513)		Rapidly Declining (Time 1: n = 104)		Decreasing Low Performance (Time 1: n = 34)	
	Diagnosed	Survived	Diagnosed	Survived	Diagnosed	Survived
	(% of survived)		(% of survived)		(% of survived)	
Time 1	16 (3%)	-	21 (20%)	-	23 (68%)	-
Time 2	29 (6%)	460 (90%)	42 (48%)	87 (84%)	18 (72%)	25 (74%)
Time 3	36 (10%)	353 (69%)	35 (66%)	53 (51%)	7 (58%)	12 (35%)
Time 4	38 (14%)	273 (53%)	15 (54%)	28 (27%)	4 (67%)	6 (18%)
Time 5	26 (13%)	206 (40%)	8 (57%)	14 (13%)	3 (75%)	4 (12%)
Ever	102 (20%)	40%	59 (57%)	13%	25(74%)	12%

Posterior Probabilities

The posterior probabilities indicate the certainty with which individuals are assigned to each class, which gives an indication of the quality of the model solution. In the three-class conditional quadratic growth mixture model, 90% of individuals (589 out of 651) were classified by a posterior probability over 0.80. However, not all the classes had the same proportion of participants classified

with over 80% certainty. In the High Performing Late Decline class, 481 out of 513 (94%), in the Rapid Decline class 75 out of 104 (72%), in the Decreasing Low Performance class 33 out of 34 (97%) were assigned to the class with 0.80 or greater level of certainty. Only 27 participants in the study were classified by posterior probabilities of less than 0.65. Of these, 13 had posterior probability of 0.50 or more of being in the High Performing Late Decline class (and so were ultimately classified as High Performing Late Decline), and the other 14 had a posterior probability of 0.50 or more of being in the Rapidly Declining class (so were classified as Rapidly Declining). The 14 individuals classified as being part of the Rapidly Declining class by a posterior probability of 0.65 or less all had a second highest probability being in the High Performance Late Decline class. Four of these 14 individuals were diagnosed with dementia at some point in the study. All individuals grouped into the High Performance Late Decline class by posterior probabilities of 0.65 or less had the second highest probability of being in the Rapidly Declining. Six were ever diagnosed with dementia.

In the unconditional three-class model with a quadratic term 91% (590 out of 651) of individuals were classified with a greater than 80% certainty. Similarly to the model with covariates, 94% of the individuals classified as Decreasing Low Performance, approximately 74% of individuals ultimately classified as Rapidly Declining, and 93% of High Performing Late Decline individuals were assigned with greater than 0.80 probabilities. Thirty-six individuals were assigned based on less than 0.65 posterior probability. The majority of the 36 (19/36) were ultimately assigned to the High Performance Late Decline class, while 16 of those remaining were assigned to the Rapid Decline class and one was assigned to the Decreasing Low Performance class. Ultimate assignments were based on posterior probabilities of greater than 0.50 in all but one case of an individual ultimately assigned to the Decreasing Low Performance class based on a 0.45 posterior probability. Examining the posterior probabilities of the cases assigned by lower levels of certainty (between 0.50 and 0.65) those ultimately classified in the High Performing Late Decline class all were next most likely to be in the Rapidly

Declining class. This means that the difficulty was in determining whether those individual's trajectories fit best with the class most likely to remain dementia free or with the Rapidly Declining class where 63% of individuals would, at some point in the study, be diagnosed with dementia. The majority of these individuals (14) were not diagnosed with dementia while in the study, indicating that assignment to the High Performance Late Decline class was likely more appropriate. In contrast those ultimately included as part of the Rapid Decline class by posterior probabilities of less than 0.65 had second highest posterior probabilities in both the High Performance Late Decline and the Decreasing Low Performance classes. Six out of the 16 individuals were diagnosed with dementia over the course of the study suggesting that most of these individuals would have been more accurately classified as belonging to the High Performing Late Decline Class.

Examining the certainty of the assignments of individuals to classes supports the conditional three-class quadratic model over the unconditional three-class quadratic model. The inclusion of the covariate (sex, age, and years of education) effects on class assignment does slightly improve the classification certainty and thus should be retained.

Analysis of subgroup with complete data until at least assessment wave 4

Given the differences in survival rates across classes, a further set of models including only participants who were in the study until at least the fourth wave ($N = 303$) was examined. The same model building strategy was attempted, however models that allowed variance around the intercept and growth parameters evidenced a variety of model convergence issues which are frequently encountered in growth mixture modeling (Jung & Wickrama, 2008). Therefore, a series of latent class growth models (LCGM), in which the variances of the intercept, slope, and quadratic terms are fixed to zero, were run. The quadratic term was included because it was significant for two out of the three latent classes (See Table 14). The purpose of this subgroup analysis was to explore whether the “best” model may have been influenced by participant attrition. The model building process was the same as that followed

previously for the full participant group. First a series of models without covariates was examined followed by a second set of LCGM's including the effects of sex, age, and years of education on class membership. Comparing the fit indices of two, three, and four class unconditional latent class growth models led to the same conclusions as those for the full sample models: the three-class model remained superior to the two-class model according to the BIC (7895.315 vs. 8212.217) but not by the VLMR-LRT. The BIC of the 4-class model was lower, but the difference was not as large as that between the two and three class models and the VLMR-LRT was not significant. Fit indices for the unconditional model are presented in Table 12. Fit indices for the conditional model are presented in Table 13 and similarly suggest the three-class conditional model is superior to the two-class model in both available fit indices. The four-class model in both model series has a lower BIC but the VLMR-LRT is not significant. The last consideration was, for both the unconditional and conditional models, that although some of the fit indices were better for the four class model, the smallest class included only 6 individuals. Therefore, the three-class model remained the better model when only participants with data until wave 4 were included. The BLRT was not considered for either the unconditional or conditional model series because the bootstrapped likelihood ratio test could not be calculated. The parameter estimates of the three classes suggest that they are similar in meaning to the classes in the three-class models with the full data set. Parameter estimates of the unconditional three-class model are shown in Table 14.

Although a second set of LCGM's including the effects of sex, age, and years of education on class membership was run (see Table 15) the inclusion of covariates resulted in nearly identical parameter estimates and identical classification so only the classification of the unconditional model is presented. The number of individuals diagnosed with dementia at each wave, within each class identified in the LCGM, was examined. Individuals were classified as belonging to a particular class based on the class for which they had the highest posterior probability, as with the previous models.

Similar labels to those given to the full sample three-class models appeared to fit the three classes in the complete data to wave 4 models.

Table 12. Fit indices for the unconditional LCGM of participants in study until at least wave 4.

Model	# of free parameters	BIC	VLMR-LRT	BLRT	Entropy	Smallest class(n)
2-class	12	8212.217	$p = .0275$	$p < .0001$	0.98	40
3-class	16	7895.315	$p = .0826$	n/a	0.95	32
4-class	20	7695.419	$p = .1578$	n/a	0.97	6

Table 13. Fit indices for the conditional LCGM of participants in study until at least wave 4.

Model	# of free parameters	BIC	VLMR-LRT	BLRT	Entropy	Smallest class(n)
2-class	15	8209.784	$p = .0657$	$p < .0001$	0.98	41
3-class	22	7889.287	$p = .0468$	n/a	0.95	32
4-class	29	7705.727	$p = .1634$	n/a	0.97	6

Table 14. Parameter estimates of three-class unconditional LGCM (restricted sample).

	Three-class model					
	High Performance Late Decline N = 204		Declining Performance N = 67		Decreasing Lower Performance N = 32	
	b	SE	b	SE	b	SE
Fixed Effects						
Intercept	28.18**	(0.12)	25.69**	(0.43)	22.16**	(1.22)
Time	0.44**	(0.07)	0.08	(0.22)	-3.17**	(0.58)
Time ²	-0.09**	(0.01)	-0.21**	(0.03)	0.07	(0.07)
Residual Variance			b	SE		
MMSE 1			8.24**	(1.98)		
MMSE 2			10.19**	(2.11)		
MMSE 3			8.36**	(1.00)		
MMSE 4			9.74**	(1.26)		
MMSE 5			17.18**	(2.26)		

* $p < .05$; ** $p \leq .001$

Table 15. Parameter estimates of the three-class conditional LGCM (restricted sample).

	High Performance Late Decline N = 203		Three-class model Declining Performance N = 68		Decreasing Lower Performance N = 32	
	b	SE	b	SE	b	SE
	Fixed Effects					
Intercept	28.17**	(0.12)	25.78**	(0.43)	22.16**	(1.21)
Time	0.44**	(0.07)	0.09	(0.22)	-3.15**	(0.57)
Time ²	-0.09**	(0.01)	-0.21**	(0.03)	0.07	(0.07)
Residual Variance			b	SE		
MMSE 1			8.26**	(1.97)		
MMSE 2			10.17**	(2.06)		
MMSE 3			8.41**	(1.03)		
MMSE 4			9.80**	(1.28)		
MMSE 5			16.90**	(2.31)		

* $p < .05$; ** $p \leq .001$ **Table 16. Three-class unconditional quadratic latent class growth model with diagnosis of dementia at each wave by class (restricted sample).**

	High Performing Late Decline N = 204	Declining Performance N = 67	Decreasing Lower Performance N = 32
Diagnosed			
Time 1	0	1 (1%)	9 (28%)
Time 2	0	4 (6%)	18 (56%)
Time 3	0	16 (24%)	26 (81%)
Time 4	2 (1%)	29 (43%)	26 (81%)
Time 5	4 (2%)	18 (51%)	14 (88%)
Ever	18(9%)	37(55%)	27(84%)

The prevalence of dementia within in each of the three classes of the final LCGM was examined. The number of people diagnosed at each wave is presented in Table 16. The percentage of individuals diagnosed at time 5, in Table 16, is calculated as the number of individuals with dementia

at time 5 out of the total number of individuals still in study at time 5. In the High Performing Late Decline class 168 of the individuals 204 (82%) who completed four assessment waves also completed the fifth. In the Declining Performance class 35 out of the 67 (52%) who completed four assessments also completed the fifth assessment, and in the Decreasing Lower Performance class 16 individuals out of the 32 (50%) in the class, who completed four assessments also completed the fifth. Individuals in the class with higher initial MMSE scores and shallower trajectories of decline were the least likely to have a diagnosis of dementia at the study outset, 0 out of 204, and only 2% were diagnosed by time 5, although an additional 14 were diagnosed on follow up. Individuals in the class with a lower initial MMSE score (Est. = 25.69), and greater negative curvature (Est. = -0.21), while unlikely to have dementia at the first wave of the study (1 out of 67), were more likely to have been diagnosed with dementia at some point (55%). In the class with the lowest initial MMSE scores, and greatest instantaneous rate of change at intercept, 9 of the 32 had dementia at study onset and 26 of 32 (81%) were diagnosed with dementia by the third wave of data collection approximately 6 years later.

Discussion

The present study investigated the use of group-based trajectory modeling for identifying cases of dementia by longitudinal MMSE scores and a small number of covariates. Overall, a three-class model, with or without covariates, was the best fit and most useful in terms of classification accuracy. Ultimately the model with covariates appeared better because of clearer classification of individuals based on posterior probabilities. Whether the best of the trajectory models is more useful for identifying cases of dementia in longitudinal studies than simpler methods likely depends on the specific concerns of the researcher.

Using a simple cut-off score of less than 24 out of 30 points on the MMSE to indicate possible dementia resulted in reasonable sensitivity (80%) similar to that reported in previous investigations (Tombaugh & McIntyre, 1992). Specificity was also reasonable (85%) and within the range of previously reported specificities. However, there were a large number of people identified by the cut-off as having dementia who did not, indicating that this method would miss-classify a substantial number of people with low MMSE scores but no dementia. For the current sample, the positive predictive value was 36%, indicating that, given an MMSE score below 24, the odds the individual does not have dementia is far greater than the odds that they do. This may not be particularly problematic in a clinical setting where further testing would presumably be done. It may be problematic in longitudinal studies where having large numbers of participants complete extensive assessments may not be feasible, particularly when a large portion of them are actually cognitively normal older adults. Excluding all individuals scoring less than the cut-off would result in an unduly selective sample. Excluding lower scoring individuals may systematically bias the sample towards younger individuals with higher levels of education (Crum, et al., 1993).

Given the relationship between age, years of education, and MMSE scores (Crum, et al., 1993), we had hypothesized that taking these factors into account might improve the classification accuracy

when considered in addition to MMSE scores. In a logistic regression analysis the covariates sex, age, and years of education did not significantly aid in diagnostic classification. Logistic regression predicts the likelihood of an outcome (in this case dementia) from a group of predictors and so is not the same as using simple cut-off scores. However, it does provide another comparison using information from only one time point. While the logistic regression was fairly accurate for distinguishing between those who did and did not have dementia at the start of the study, the accuracy was primarily in correctly classifying those without dementia, while less than half, 45%, of individuals with dementia were correctly identified as such. These findings confirm the limitations of simpler methods for screening with the MMSE that were suggested by the extant literature.

Using group-based trajectory modeling, with repeated MMSE scores as the dependent variable and, in some models, a selection of covariates, to identify cases of dementia by class assignment was of mixed success. The superior models, in terms of accurate classification of individuals' diagnostic status and several fit indices, were the three-class growth mixture models with linear and quadratic growth terms. The three-class model with the covariates, sex, mean centered age, and years of education, resulted in classes with nearly identical size and meaning to the three-class unconditional model. The trajectory shapes were not substantially different yet the conditional model assigned individuals to classes with greater certainty. Therefore the three-class model with quadratic change and with covariate effects on class membership is considered the "final" model.

Considering accuracy of dementia detection in terms of sensitivity, specificity, positive predictive value, and negative predictive value is less meaningful for the final model because two of the three classes need to be grouped so that predicted dementia status is a dichotomy. For comparative purposes, the classes with substantial proportions of individuals with dementia can reasonably be grouped together. When considered this way, the accuracy of the final model in identifying individuals who were ever diagnosed with dementia may be better than using 24/30 on the MMSE at time 1 or time

5 as a cut-off score only for particular purposes. The sensitivity of the growth mixture model is considerably lower than the sensitivity of using the 24 cut-off score either at time 1 or time 5 and there was a decrease in negative predictive value. However, specificity was slightly, and positive predictive value was substantially, better. Whether the trade-off of lower sensitivity is worthwhile would depend on the aims of the researcher using the model. For example, those primarily interested in ensuring that individuals with dementia are excluded from a sample may prefer other methods, but those who want to identify a population of individuals who have dementia would benefit from the improved positive predictive value. In comparing the trajectory model results to the accuracy of using the cut-off score for the fifth assessment wave the differences were similar. The trajectory model had superior specificity and positive predictive value, slightly worse negative predictive value, and nearly half the sensitivity. Neither comparison is direct however, as the trajectory models take into account up to five different MMSE scores. For the group-based trajectory models, prediction of dementia was considered accurate if the individual was classified as “Rapidly Declining” or “Decreasing Low Performance” by the model and they were ever diagnosed with dementia. In contrast, the cut-off score accuracy values were established based on comparing one MMSE score to current diagnostic status. When looking at individuals who survived in study until the last wave of assessment, accuracy of the conditional three-class quadratic growth mixture model classes in identifying who is diagnosed with dementia at or by the last wave is superior to using the 24/30 cut-off score in all but sensitivity. This is perhaps unsurprising as we have the most information about these individuals cognitive function over time.

Examining the proportion of individuals who were diagnosed with dementia at each wave of the study in each class provided some additional insight into the characteristics of the identified classes. Most of the “Decreasing Low Performance” class in the conditional three-class model were diagnosed with dementia at or by the first wave of the study. For these individuals, onset of dementia may have been recent or much earlier. At each assessment wave, only a few surviving individuals in the

Decreasing Low Performance class did not have dementia. Given that the majority of attrition was due to death, it is plausible that the nine individuals in the class that remained free of dementia performed poorly on the MMSE due to failing health. The second, “Rapidly Declining”, class was comprised of individuals, the majority of whom were not diagnosed with dementia at the beginning of the study, but were likely to be diagnosed over the study period. By or at the third wave of assessment, the majority (66%) of participants still in study were diagnosed with dementia; even with participant dropout, those diagnosed with dementia remained a majority for the fourth and fifth assessments. However, by the last assessment only 13% of participants originally assigned to the Rapidly Declining class, and 12% of those in the Decreasing Low Performance class, were still in study. The “High Performance Late Decline” class had the highest rate of survival in study by a large margin, with 40% surviving to the last wave of assessment.

There are several possibilities for why survival rates varied by class. It is possible that the model is substantially influenced by participant attrition. In a simulation study, participant attrition substantially affected group size, formed based on posterior probabilities, and it was particularly problematic when groups are not well separated initially (Haviland, Jones, & Nagin, 2011). The High Performance Late Decline, Rapidly Declining, and Decreasing Low Performance groups were well separated but this does not rule out the possibility that attrition influenced the model results. In the current study, this possibility was tested by employing the same modelling strategy on a subsample of participants, those who remained in study for at least four waves. The same model building strategy was attempted so that the results would be directly comparable, however, the three-class growth mixture model had model convergence issues. Instead a latent class growth model was developed, which differs from growth mixture models in that, in the former, variance within classes is fixed to zero. This assumes that all variance in the sample is captured by the different growth parameters of multiple latent classes, an assumption which may or may not be a valid. However, examining growth

trajectories for a subset of individuals in each class supported the tenability of the zero variance assumption. The similarities between the best latent class growth model, with the smaller more select sample, and the best growth mixture model with the full sample, supports the validity of the three trajectory class models and suggests that the final model based on the full dataset was not grossly biased by attrition.

However, several differences between the growth mixture model based on the full data set and the latent class growth model based on the selected sample were evident. Namely, for each class the mean initial MMSE scores for the sample of those with at least four waves of data were higher than those of the full sample. This difference was particularly large for the Decreasing Low Performance group where the intercept was 22, compared to 10 for the full sample. This difference is likely because individuals who had been living with dementia prior to study enrollment or early in the study period were less likely to survive until the fourth wave. The average life expectancy following diagnosis of Alzheimer disease depends on age at diagnosis, with an average of just over 8 years for a person diagnosed at 65 to just over 3 years for someone diagnosed at 90 (Brookmeyer et al., 2002). Given that participants were at least 80 years old at baseline, few individuals diagnosed prior to, or at the start of the study would have survived to the fourth assessment six years later. It is also possible that individuals whose lower MMSE scores were related to other health conditions would have been similarly less likely to survive until the fourth wave. Despite these differences, the general trajectory shapes of the latent class growth model classes with the subsample were fairly consistent with those of previous models. In the model based on the restricted sample, there was a class characterized by high MMSE scores with a late decline in performance, evidenced by a significant quadratic term. This is similar to the High Performance Late Decline class of the three-class models with the full sample. The second class with the restricted sample is characterized by a lower intercept but a larger significant quadratic term suggesting greater decline. This can be seen as analogous to the Rapid Decline class of

three-class model with the full sample in that the class captures a group of individuals in between the other two classes. The last class of the three-class model LCGM with the restricted sample has a larger significant linear slope term, which indicates significant decline at intercept, suggesting a more immediate decline in performance that might characterize individuals who already have dementia. However, the three-class quadratic latent growth curve model was not substantially better at predicting dementia status than the three-class models with the full sample. This suggests that the lack of predictive accuracy of the group-based trajectory models is likely not due to the effects of attrition.

In dementia due to Alzheimer's disease, cognitive decline has been found to begin at least three years prior to a diagnosis (Bäckman, Jones, Berger, Laukka, & Small, 2005). This suggests that the three-class trajectory models may be correctly identifying individuals whose cognitive decline is due to the disease process but who were not diagnosed before death or leaving the study. It is also important to consider that diagnostic processes for dementia, although used as the "gold standard" in this study, do not have perfect accuracy themselves when compared to other methods such as post-mortem autopsy (McKhann et al., 2011). Thus, it is possible that some of the mismatch between the group-based trajectory model class predictions of dementia status and diagnosed dementia status is due to inaccuracy in the diagnostic process, rather than the model predictions. In particular, the diagnostic criteria used allows only for individuals to have dementia or not, but some theories suggest that dementia may actually be an extreme on a continuum of cognitive aging (Walters, 2010).

Another, not unrelated, possibility is that some individuals may be experiencing cognitive decline that is related to other health concerns and not a reflection of dementia. Substantial evidence supports the association between many aspects of health and cognitive function (Spiro & Brady, 2011). Many conditions that increase in prevalence among older adults, for example, low blood pressure and other cardiovascular health issues, are associated with declines in cognitive function not considered

dementia (Bossola, et al., 2011; Molander, et al., 2010). Attempts to account for the role of additional health factors in identifying cases of dementia may be a direction for future research.

The initial hypothesis was that part of the difficulty of screening individuals based on a single time point is that individuals who have always been lower or higher functioning than average may be more likely to be falsely classified (Cummings, 1993; Shiroky et al., 2007). Given that a decline from previous levels of functioning is part of the diagnostic criteria for dementia, accurate knowledge of an individual's previous ability should improve diagnostic accuracy. Grouping individuals by longitudinal trajectories should help this issue, as within person change over time is considered rather than just individual differences. Thus individuals should be better characterized by whether or not their cognitive ability is declining in comparison to their own baseline, than by absolute scores on cognitive measures. As expected, previous studies have shown that individuals with dementia evidence different patterns of cognitive function over time than older adults experiencing normative aging (Small & Bäckman, 2007). The results of the present study further support this conclusion, as the models with the "best" comparative fit were the three-class models, with or without covariates, confirming that individuals differed in their trajectory of MMSE scores over time.

Further, we had hypothesized that considering covariates that have been previously related to MMSE performance, even without longitudinal information, would improve diagnostic accuracy. This was not supported by the logistic regression analysis. Although the three-class model with covariate effects on class membership was ultimately preferred, the average trajectory of MMSE scores in each of the three classes was very similar to the three-class model without covariates. The size and meaning of the classes was not largely affected by the inclusion of covariate effects on class, confirming minimal effects.

However, two out of the three covariates were significantly related to the odds of being assigned to a particular class in ways that are consistent with previous research. In the final three-class

model, sex did not significantly impact class membership for any of the comparisons. It is, perhaps, unsurprising given that evidence for a sex effect in the development of dementia is somewhat mixed. Level of education, in contrast, has been shown to relate to dementia rates, such that individuals with more education are less likely to receive a diagnosis of dementia (Crum, et al., 1993; Dufouil, et al., 2000). This was reflected in our model; individuals with greater years of education were more likely to be classified in the High Performance Late Decline class, which had lower rates of dementia, as opposed to the Decreasing Low Performance class which had higher rates of dementia. Rates of dementia have also been found to increase with increased age (Gao et al., 1998). For our model, individuals were more likely to be in the Rapidly Declining class than the High Performance Late Decline class if they were older, but the comparisons between other classes were not significant. This may, in part, reflect the very restricted age range of participants in the study, as all must have been 80 years old or older at study start.

Two previous studies used similar methodology, with MMSE scores over time as the dependent variable. Small and Bäckman (2007) conducted a study using growth mixture modeling to identify those who would go on to develop dementia out of a sample of older adults. They, however, tested only a one-class versus two-class model and found the two-class model superior. In their final model, one class appeared to be those who remained dementia free over the follow-up period and the other appeared to be composed of those who would receive a diagnosis of Alzheimer's disease. They tested the sensitivity and specificity of their two-class model in the same way as was done in the present study and found much better classification accuracy than was found in this analysis. In Small and Bäckman (2007), 68% of preclinical AD cases and 86% of those who would remain dementia free were correctly identified. The current two-class model correctly identified only 15% of dementia cases. This is likely due to differences in sample, as the current study included participants with a greater range of cognitive statuses, including those who would remain dementia free, those who would develop dementia over the

study period, and those who had previously been diagnosed. The low intercept, 11 points, and the shape of the trajectories of our two-class model suggests that the model is separating those who have moderate to severe dementia from the rest of the sample, but not differentiating among the others. Despite these differences, the High Performance Late Decline trajectory was similar to the class of individuals who would remain dementia free in Small and Bäckman (2007). In their study, the class of individuals who would remain free of dementia had a mean initial MMSE score of 27.89, and also had a non-significant linear slope, and a significant quadratic slope term. This suggests that the results of the present analysis may be generalizable to other older adult populations.

Also using similar methodology, Wilkosz et al. (2010) found six latent trajectory groups in their analysis focussing on the course of cognitive decline in Alzheimer's disease. They included only individuals with a diagnosis of possible or probable Alzheimer's disease who completed the MMSE each year for a minimum of four years. Smaller testing intervals may allow for a more fine-grained examination of trajectories of change and thus more trajectory groups. They do acknowledge, however, that examining plots of the six trajectory groups suggests that the groups could have been subsumed into two. The authors explain that they did not do this because statistical fit indices were superior for the six-class model. The population in Wilkosz et al. (2010) included only those diagnosed with Alzheimer's disease, the majority of whom were in the mild stage, which might suggest they should have found fewer classes, not more than were found in the current study. However, their participants were more varied in terms of demographic characteristics, which, they found, were related to trajectory group assignment. Participants between 60 and 70 years old were more likely to be assigned to the class that showed the most rapid decline, which corresponds to other research suggesting those who suffer from Alzheimer's disease at younger ages have a greater burden of plaques and tangles and, once symptoms are present, show a rapid decline. Given that all participants in the current study were 80 years old at the first assessment, those who developed dementia in their 60's would likely have not

been included due to inability to complete even the most basic testing or having already died. The cases of dementia in our sample may, therefore, be more homogenous than those in Wilkosz et al. (2010). It may also be a reflection of the model specification used, as latent class growth models often result in more classes than growth mixture models because, in the former, all variance must be accounted for by different classes rather than variance within classes (Tofighi & Enders, 2008).

Implications

The current findings have several implications for researchers using longitudinal data sets where time and resources for establishing diagnosis through more established methods are not appropriate or feasible. In this sample, group-based trajectory models do not yield significant benefits in terms of classification accuracy, when a three-class model is interpreted as a dichotomy, those with dementia and those without. However, this is unsurprising as collapsing two classes into one, likely ignores the very differences that make the three-class models superior to the two-class models. The data was best characterized by the conditional and unconditional three class models, but the utility of these models depends on the question being asked. The three-class models were all characterized by a class in which individuals had a high likelihood of remaining dementia free, ranging from a greater than 90% chance in the model with complete data until wave four, to a just less than 85% chance in the whole sample three-class model with covariates. This is, at least, about as accurate as the simple greater than 24/30 cut-off used to screen out cognitively impaired individuals. Considering only the class representing those most likely to develop dementia, the “Decreasing Low Performance” class, the probability that an individual assigned to that class actually has or will develop dementia is 74% or higher in all of the three-class models. Thinking of this as a positive predictive value, it is substantially better than the traditional 24 cut-off but not quite as accurate as logistic regression. However, logistic regression can be used only when the outcome, dementia status, is known. Including the outcome in the analysis is likely improving the accuracy. Outcomes can also be specified within growth mixture and

latent class growth models but this was not done in the present analysis because the purpose was to develop a method that could be used to identify dementia cases in samples where diagnosis is not known. If outcomes had been included in the group-based trajectory modeling the outcome influences would have influenced the formation of classes and limited the generalizability to studies without diagnosis (Petras & Maysn, 2011). It is the “Rapidly Declining” class, those individuals with a mean initial MMSE score that falls in between the other two classes, that are the most difficult to accurately classify. However, the odds of an individual assigned to that class having or later receiving a diagnosis of dementia is still greater than 50%. It is possible that the difficulty in classifying the “Rapidly Declining” group is due to characteristics of the individuals themselves and not necessarily limitations of the model.

That different classes offer different degrees of certainty is not necessarily problematic as it may allow researchers different degrees of caution in which participants they exclude and include in any given analysis. If one is primarily interested in investigating a cognitively normative population then it may be best to include only individuals assigned to the first class. However, in studies where the group of interest is those with dementia, including both the second and third classes may be desirable, particularly if there is an opportunity to administer additional cognitive measures.

Using cut-off scores to distinguish between groups using the MMSE is less effective at separating mild dementia from normative aging than severe dementia from normative aging. The “Rapidly Declining” class in the three-class models somewhat addresses this issue by allowing such individuals to be separately identified. It is also important to reiterate that even a more extensive diagnostic process is not 100% accurate, and this may be a further source of mismatches between the trajectory class prediction of dementia, and “actual” diagnosis. Mild dementia cases are, likely, more difficult to diagnose even using more extensive cognitive testing.

Limitations and future directions

While the current study outlines a possible approach for improved grouping of older adult participants according to dementia status, it does so with several limitations. First, the use of a single measure, while purposely selected here due to its popularity, is inherently limited. The limitations of the measure were discussed, but more advanced analytical methods are unlikely to transform a screening tool into a diagnostic test.

A second limitation lies with the analytical approach chosen. Group-based trajectory modeling, growth mixture modeling in particular, is a relatively new technique, and as such the information on factors that might lead to spurious conclusions are not fully understood. The current literature was considered and guidelines followed, but it should be acknowledged that the enumeration of classes and selection of models is subjective (Bauer & Curran, 2003). As Muthén (2003) points out, this does not invalidate the method, but of course, suggests that there is considerable room for interpretation. Other decisions, such as fixing the variances across classes, were made to achieve a result, but it is questionable whether equal variances across classes actually represent reality. It is unclear what effect factors such as this have on parameter estimates and the size and meaning of classes.

Although touched on previously, it is important to reiterate that the participants in this study were a very select group of the oldest old twin pairs in Sweden, and thus generalizability to other samples is uncertain. Individuals who survive to their eighth or ninth decade of life may represent a select group, who may have particular lifestyles or genetic predispositions that are different from those of individuals who do not survive as long. A logistic regression analysis allowing the lack of independence of twin-pairs to be controlled suggested it had little effect and was subsequently ignored, but it is possible the inclusion of twin pairs affected model outcomes. Extending the present findings to other studies will be the true test of whether the use of group-based trajectory modeling for the identification of dementia is beneficial.

The central goal of the present analysis was to investigate the feasibility and usefulness of group-based trajectory modeling techniques for identifying cases of dementia in longitudinal datasets. Future studies are needed to examine whether the conditional three-class quadratic model that was considered the best model for this sample can also characterize other samples. Examining the classification accuracy of the model in a population with a wider age range is particularly important. The final model did improve the accuracy of identifying those without dementia, and individuals who were assigned to either the Rapidly Declining or Decreasing Low Performance groups were likely to have, or develop, dementia. Other longitudinal studies with diagnoses are needed to establish whether using group-based trajectory modeling in longitudinal datasets without a full diagnosis may be useful. Current findings suggest group-based trajectory models may at least be used to efficiently identify individuals in need of further assessment, which may be particularly helpful for ongoing studies.

References

Abbott, A. (2011). Dementia: A problem for our age. *Nature*, 475(7355), S2-S4.

Alzheimer Society of Canada. (2010). *Rising tide: The impact of dementia on Canadian Society*.

Retrieved from <http://www.alzheimer.ca/en/Get-involved/Raise-your-voice/Rising-Tide/Rising-tide-summary>.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Anderson, T. M., Sachdev, P. S., Brodaty, H., Trollor, J. N., & Andrews, G. (2007). Effects of sociodemographic and health variables on Mini-Mental State Exam scores in older Australians. *American Journal of Geriatric Psychiatry*, 15(6), 467-476. doi: 10.1097/JGP.0b013e3180547053

Bäckman, L., Jones, S., Berger, A. K., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*, 19(4), 520-531. doi: 10.1037/0894-4105.19.4.520

Barnes, D. E., & Yaffe, K. (2011). The projected effect of risk factor reduction on Alzheimer's disease prevalence. *The Lancet Neurology*, 10(9), 819-828. doi: 10.1016/s1474-4422(11)70072-2

- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8(3), 338-363. doi: 10.1037/1082-989x.8.3.338
- Bossola, M., Antocicco, M., Di Stasio, E., Ciciarelli, C., Luciani, G., Tazza, L., . . . Onder, G. (2011). Mini Mental State Examination over time in chronic hemodialysis patients. *Journal of Psychosomatic Research*, 71(1), 50-54. doi: 10.1016/j.jpsychores.2011.01.001
- Brookmeyer, R. & Curriero, F. C. (2002). Survival curve estimation with partial non-random exposure information. *Statistics in Medicine*, 21(18), 2671-2683.
- Brungnolo, A., Nobili, F., Barbieri, M. P., Dessi, B., Ferro, A., Girtler, N., . . . Rodriguez, G. (2009). The factorial structure of the Mini Mental State Examination in Alzheimer's disease. *Archives of Gerontology and Geriatrics*, 49, 180-185.
- Burton, C. L., Strauss, E., Bunce, D., Hunter, M. A., & Hultsch, D. F. (2009). Functional abilities in older adults with mild cognitive impairment. *Gerontology*, 55(5), 570.
- Government of Canada. (2012). *Jobs, growth, and long-term prosperity: Economic action plan 2012*. Retrieved from <http://www.budget.gc.ca/2012/plan/toc-tdm-eng.html>.
- Cacho, J., Benito-León, J., Garcia-Garcia, R., Fernandez-Calvo, B., Vicente-Villardón, J. L., & Mitchell, A. L. (2010). Does the combination of the MMSE and clock drawing test (mini-clock) improve the detection of mild Alzheimer's disease and mild cognitive impairment? *Journal of Alzheimer's Disease*, 22(3), 889-896.

- Christensen, H., Mackinnon, A., Jorm, A. F., Korten, A., Jacomb, P., Hofer, S. M., & Henderson, S. (2004). The Canberra Longitudinal Study: Design, aims, methodology, outcomes, and recent empirical investigations. *Aging, Neuropsychology, and Cognition, 11*, 169-195.
- Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: Predictors and consequences. *Journal of Consulting and Clinical Psychology, 70*(4), 976-985.
- Crum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and education level. *Journal of the American Medical Association, 269*(18), 2386-2391. doi: 10.1001/jama.269.18.2386
- Cummings, J. L. (1993). Mini-Mental State Examination: Norms, normals, and numbers. *The Journal of the American Medical Association, 269*(18), 2420-2421.
- Davey, R. J., & Jamieson, S. (2004). The validity of using the mini mental state examination in NICE dementia guidelines. *Journal of Neurology Neurosurgery and Psychiatry, 75*(2), 343-344.
- Davidson, J. E. Irizarry, M. C., Bray, B. C., Wetten, S., Galwey, N. Gibson, R.,...Monsch, A. U. (2010). An exploration of cognitive subgroups in Alzheimer's disease. *Journal of the International Neuropsychological Society, 16*(2), 233-243.
- Daviglius, M. L., Plassman, B. L., Pirzada, A., Bell, C. C., Bowen, P. E., Burke, J. R., . . . Williams, J. W. (2011). Risk Factors and Preventive Interventions for Alzheimer Disease State of the Science. *Archives of Neurology, 68*(9), 1185-1190. doi: 10.1001/archneurol.2011.100
- Dufouil, C., Clayton, D., Brayne, C., Chi, L. Y., Dening, T. R., Paykel, E. S., . . . Huppert, F. A. (2000). Population norms for the MMSE in the very old. *Neurology, 55*, 1609-1613.

- Dureman, I., & Sälde, H. (1959). *Psychometric and experimental methods for the clinical evaluation of mental functioning*. Stockholm.
- Eisenstein, M. (2011). Genetics: Finding risk factors. *Nature*, *475*(7355), S20-S22.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189-198. doi: 10.1016/0022-3956(75)90026-6
- Gao, S., Hendrie, H. C., Hall, K. S., & Hui, S. (1998). The relationships between age, sex and the incidence of dementia and Alzheimer disease: A meta-analysis. *Archives of General Psychiatry*, *55*(9), 809-815.
- Grober, E., Hall, C., Lipton, R. B., & Teresi, J. A. (2008). Primary care screen for early dementia. *Journal of American Geriatric Society*, *56*, 206-213.
- Grut, M., Fratiglioni, L., Miitonen, M., & Winblad, B. (1993). Accuracy of the Mini-Mental Status Examination as a screening test for dementia in a Swedish elderly population. *Acta Neurologica Scandinavica*, *87*, 312-317.
- Guo, Z. C., Fratiglioni, L., Winblad, B., & Viitonen, M. (1997). Blood pressure and performance on the Mini-Mental State Examination in the very old - Cross-sectional and longitudinal data from the Kungsholmen Project. *American Journal of Epidemiology*, *145*(12), 1106-1113.
- Haviland, A. M., Jones, B. L., & Nagin, D. S. (2011). Group-based Trajectory Modeling Extended to Account for Nonrandom Participant Attrition. *Sociological Methods & Research*, *40*(2), 367-390. doi: 10.1177/0049124111400041

- Hofer, S. M., & Piccinin, A. (2007). Longitudinal studies of aging. In J. E. Birren (Ed.), *Encyclopedia of gerontology* (2nd ed.). Oxford: Elsevier.
- Hofer, S. M., & Sliwinski, M. J. (2001). Understanding ageing: An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology*(47), 341-352.
- Huang, D., Brecht, M. L., Hara, M., & Hser, Y. I. (2010). Influences of covariates on growth mixture modeling. *Journal of Drug Issues*, 40(1), 173-194.
- Huff, F. J., Becker, J. T., Belle, S. H., Nebes, R. D., Holland, A. L., & Boller, F. (1987). Cognitive deficits and clinical diagnosis of Alzheimer's disease. *Neurology*, 37, 1119-1124.
- Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D., & Dartiques, J. (1997). A 5-year longitudinal study of the Mini-Mental State Examination in normal aging. *American Journal of Epidemiology*, 145(6), 498-506.
- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2/1, 302-317.
- Leoutsakos, J. M., Muthén, B., Breitner, J. C. S., Lyketsos, C. G., & Team., A. R. (2012). Effects of non-steroidal anti-inflammatory drug treatments on cognitive decline vary by phase of pre-clinical Alzheimer disease: findings from the randomized controlled Alzheimer's Disease Anti-inflammatory Prevention Trial. *International Journal of Geriatric Psychiatry*, 27, 364-374.
- Lo, Y. T., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778. doi: 10.1093/biomet/88.3.767

- Manly, J. J., Schupf, N., Stern, Y., Brickman, A. M., Tang, M. X., & Mayeux, R. (2011). Telephone-Based Identification of Mild Cognitive Impairment and Dementia in a Multicultural Cohort. *Archives of Neurology*, *68*(5), 607-614.
- McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Petrill, S. A., & Plomin, R. (1997). Substantial Genetic Influence on Cognitive Abilities in Twins 80 or More Years Old. *Science*, *276*, 1560-1563.
- McDowell, I., Hill, G., & Lindsay, J. (2001). An overview of the Canadian Study of Health and Aging. *International Psychogeriatrics*(13(Suppl 1)), 7-18.
- McKhann, G., Drachman, D., Folstein, M. F., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease. *Neurology*, *34*(7), 939.
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers & Dementia*, *7*(3), 263-269. doi: 10.1016/j.jalz.2011.03.005
- Mitchell, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research*, *43*, 411-431.
- Molander, L., Gustafson, Y., & Lovheim, H. (2010). Low Blood Pressure Is Associated with Cognitive Impairment in Very Old People. *Dementia and Geriatric Cognitive Disorders*, *29*(4), 335-341. doi: 10.1159/000289821

- Moraes, C., Pinto, J. A., Lopes, M. A., Litvoc, J., & Bottino, C. M. C. (2010). Impact of sociodemographic and health variables on mini-mental state examination in a community-based sample of older people. *European Archives of Psychiatry and Clinical Neuroscience*, 260(7), 535-542. doi: 10.1007/s00406-010-0104-3
- Morris, J. C. (1993). The Clinical Dementia Rating Scale (CDR); current version and scoring rules. *Neurology*, 43, 2412-2414.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences*. Newbury Park: Sage.
- Muthén, B., & Asparoutiov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs (Eds.), *Longitudinal Data Analysis* (pp. 143-165). Boca Raton: Chapman & Hall.
- Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., . . . Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, 3(4), 459-475. doi: 10.1093/biostatistics/3.4.459
- Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463-469.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4(2), 139-157.
- Nagin, D. S., & Land, K. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, 31, 327-362.

- Nagin, D. S., & Tremblay, R. E. (2005). Developmental trajectory groups: Fact or a useful statistical fiction? *Criminology*, *43*(4), 873-904. doi: 10.1111/j.1745-9125.2005.00026.x
- Nilsson, F. M. (2007). Mini Mental State Examination (MMSE) - probably one of the most cited papers in health science. *Acta Neurologica Scandinavica*, *116*(2), 156-157. doi: 10.1111/j.1600-0447.2007.01037.x
- Nylund, K. L., Asparoutiov, T., & Muthen, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling-a Multidisciplinary Journal*, *14*(4), 535-569.
- Olazaran, J., Trincado, R., Bermejo, F., Benito-Leon, J., Diaz, J., & Vega, S. (2004). Selective memory impairment on an adapted Mini-Mental State Examination increases risk of future dementia. *International Journal of Geriatric Psychiatry*, *19*(12), 1173-1180. doi: 10.1002/gps.1236
- Park, D., & Reuter-Lorenz, P. A. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology*, *60*, 17-196.
- Pedersen, N. L., Gatz, M., Berg, S., & Johansson, B. (2004). How heritable is Alzheimer's disease late in life? Findings from Swedish twins. *Ann Neurol*, *55*(2), 180-185. doi: 10.1002/ana.10999
- Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology*: Springer.
- Raina, P. S., Wolfson, C., Kirkland, S. A., Keshavarz, H., Griffith, L. E., Patterson, C., . . . Angus, C. L. (2009). Ascertainment of Chronic Diseases in the Canadian Longitudinal Study on Aging (CLSA), Systematic Review. *Canadian journal on aging*, *28*(03), 275.

- Rasmusson, D. X., Carson, K. A., Brookmeyer, R., Kawas, C., & Brandt, J. (1996). Predicting rate of cognitive decline in probable Alzheimer's disease. *Brain and Cognition, 31*(2), 133-147.
- Ratcliff, G., Dodge, H., Birzescu, M., & Ganguli, M. (2010). Tracking cognitive function over time: Ten-year longitudinal data from a community-based study. *Applied Neuropsychology: Adult, 10*(2), 76-88.
- Reynolds, M. D., Johnston, J. M., Dodge, H. H., DeKosky, S. T., & Ganguli, M. (1999). Small head size is related to low Mini-Mental State Examination scores in a community sample of nondemented older adults. *Neurology, 53*(1), 228-229.
- Sampson, R. J., Laub, J. H., & Eggleston, E. P. (2004). On the robustness and validity of groups. *Journal of Quantitative Criminology, 20*(1), 37-42. doi: 10.1023/b:joqc.0000016698.36239.91
- Schaeffer, C. M., Petras, H., Ialongo, N., Poduska, J., & Kellam, S. (2003). Modeling growth in boys' aggressive behavior across elementary school: links to later criminal involvement, conduct disorder, and antisocial personality disorder. *Developmental Psychology, 39*(6), 1020-1035.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of statistics, 6*(2), 461.
- Shiroky, J. S., Schipper, H. M., Bergman, H., & Chertkow, H. (2007). Can you have dementia with an MMSE score of 30?. *Alzheimer's Disease and Other Dementias, 22*(5), 406-415.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Small, B. J., & Bäckman, L. (2007). Longitudinal trajectories of cognitive change in preclinical Alzheimer's Disease: A growth mixture modeling analysis. *Cortex, 43*, 826-834.

- Small, B. J., Viitanen, M., & Bäckman, L. (1997). Mini-mental state examination item scores as predictors of Alzheimer's disease: Incidence data from the Kungsholmen Project, Stockholm. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*, 52(5), M299-M304.
- Spiro, A., & Brady, C. B. (2011). Integrating Health into Cognitive Aging: Toward a Preventive Cognitive Neuroscience of Aging. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, 66, 17-25. doi: 10.1093/geronb/gbr018
- Tofighi, D., & Enders, C. K. (2008). *Identifying the correct number of classes in growth mixture models*. Charlotte: Information Age Publishing-Iap.
- Toglia, J., Fitzgerald, K. A., O'Dell, M. W., Mastrogiovanni, A. R., & Lin, C. D. (2011). The Mini-Mental State Examination and Montreal Cognitive Assessment in Persons With Mild Subacute Stroke: Relationship to Functional Outcome. *Archives of Physical Medicine and Rehabilitation*, 92(5), 792-798. doi: 10.1016/j.apmr.2010.12.034
- Tombaugh, T. N. (2005). Test-retest reliable coefficients and 5-year change scores for the MMSE and 3MS. *Archives of Clinical Neuropsychology*, 20(4), 485-503. doi: 10.1016/j.acn.2004.11.004
- Tombaugh, T. N. & McIntyre, N. J. (1992). The Mini-Mental State Examination: A Comprehensive Review. *Journal of the American Geriatrics Society*, 40(9), 922-935.
- Tuokko, H., Morris, C., & Ebert, P. (2005). Mild cognitive impairment and everyday functioning in older adults. *Neurocase*, 11(1), 40-47. doi: 10.1080/13554790490896802
- Walters, G. D. (2010). Dementia: Continuum or distinct entity. *Psychology of Aging*, 25, 534-544.

Wilkosz, P. A., Seltman, H. J., Devlin, B., Weamer, E. A., Lopez, O. L., DeKosky, S. T., & Sweet,

R. A. (2010). Trajectories of cognitive decline in Alzheimer's disease. *International*

Psychogeriatrics, 22(2), 281-290. doi: 10.1017/s1041610209991001