Including Persistency of Impairment in Mild Cognitive Impairment Classification Enhances Prediction of 5-Year Decline

Susan Vandermorris1,2,*, David F. Hultsch1, Michael A. Hunter1, Stuart W.S. MacDonald1, Esther Strauss1

1Department of Psychology, University of Victoria, Victoria, British Columbia, Canada
2Rotman Research Institute, Toronto, Ontario, Canada

*Corresponding author at: Rotman Research Institute, Baycrest, 3560 Bathurst Street, Toronto, ON, Canada M6A 2E1. Tel.: +1-416-884-0035.
Fax: +1 416-785-2862.
E-mail address: susan.vanderhill@aya.yale.edu (S. Vandermorris).

Accepted 11 November 2010

Abstract

Although older adults with Mild Cognitive Impairment (MCI) show elevated rates of conversion to dementia as a group, heterogeneity of outcomes is common at the individual level. Using data from a prospective 5-year longitudinal investigation of cognitive change in healthy older adults (N = 262, aged 64–92 years), this study addressed limitations in contemporary MCI identification procedures which rely on single occasion assessment (“Single-Assessment [SA] MCI”) by evaluating an alternate operational definition of MCI requiring evidence of persistent cognitive impairment over multiple-testing sessions (“Multiple-Assessment [MA] MCI”). As hypothesized, prevalence of SA-MCI exceeded that of MA-MCI. Further, the MA-MCI groups showed lower baseline cognitive and functional performance and steeper cognitive decline compared with Control and SA-MCI group. Results are discussed with reference to retest effects and clinical implications.

Keywords: Mild Cognitive Impairment; Assessment; Aging; Dementia; Practice Effects

Introduction

Advances in treatments for Alzheimer’s disease and related dementia syndromes (e.g., Vascular Dementia, Dementia with Lewy Bodies, etc.) have shifted research efforts toward the early identification of individuals likely to develop these disorders. The term Mild Cognitive Impairment (MCI) has been used as a label for individuals who show cognitive impairment relative to their healthy peers, but do not meet full criteria for any dementia syndrome (Flicker, Ferris, & Reisberg, 1991; Smith, Petersen, Parisi, & Ivnik, 1996). Although MCI has been conceptualized as a precursor to dementia (Petersen et al., 1999), evidence for the utility of MCI as a predictor of impending dementia has been found to be somewhat limited. Individuals with MCI do show elevated rates of conversion to dementia at the group level, but heterogeneity of outcomes is common at the individual level (for reviews, see Bruscoli & Lovestone, 2004; Petersen, 2004; Tuokko & McDowell, 2006). Some individuals with MCI develop dementia, some remain stable for long periods, and some revert to unimpaired status. The aim of the current study was to investigate revised classification procedures to enhance existing methods for identifying those at greatest risk of dementia.

Evidence to support the utility of MCI as a predictor of dementia comes from longitudinal studies, which have produced varying prevalence and incidence rates for MCI and varying rates of conversion from MCI to dementia. A recent review of nearly 40 major studies of MCI identified prevalence estimates for MCI in older adult samples ranging from 1% to 36%, incidence rates for MCI (less frequently reported) ranging from 8 to 77 per 1,000 per year, and rates of conversion from MCI to dementia ranging from 5% to 30% per year (Tuokko & McDowell, 2006). The authors proposed that these variations may be

© The Author 2010. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
accounted for by differences in syndromes studied (e.g., Age Associated Memory Impairment vs. Aging-Associated Cognitive Decline), age of study participants, operational definitions of “impairment,” statistical methodologies, presentation of results, study design, and sampling and measurement techniques. For example, advancing age is consistently associated with increased risk for conversion from MCI to dementia (e.g., Kryscio, Schmitt, Salazar, Mendiondo, & Markesbery, 2006; Visser, Kester, Jolles, & Verhey, 2006). There is evidence that certain variations in operational definitions for MCI are associated with increased risk of decline and/or conversion to Alzheimer’s disease (e.g., criteria that require more pervasive cognitive impairment, Palmer, Backman, Winblad, & Fratiglioni, 2008; require specific impairment in the memory domain, Tuokko et al., 2003; or both, De Jager, Blackwell, Budge, & Sahakian, 2005), though relatively broad definitions of MCI, which specify impairment in any cognitive domain, have shown comparable associations with memory decline (Schonknecht, Pantel, Kruse, & Schroder, 2005) and all-cause dementia (Tuokko et al., 2003). There is also evidence that the quantification of biomarkers such as cerebrospinal fluid levels of β-amyloid, τ-protein (e.g., Mattsson et al., 2009), and neuroimaging measures, such as MRI-derived medial temporal atrophy (e.g., DeCarli et al., 2007), improve prediction accuracy in MCI, especially when used in combination (e.g., Vemuri et al., 2009).

Although variations in prevalence, incidence, and rates of conversion to dementia raise questions about the utility of the MCI classification as an indicator of impending dementia, perhaps more striking are the pervasive reports that a significant minority of individuals classified as MCI at one time point fail to demonstrate cognitive impairment at a subsequent time point. Such individuals are said to “revert” from MCI to unimpaired or cognitively normal status. Rates of reversion have been reported in a number of major epidemiological samples, ranging from 15% to 41% over 1–2-year follow-up intervals (Busse, Hensel, Guhne, Angermeyer, & Riedel-Heller, 2006; Ganguli, Dodge, Shen, & DeKosky, 2004; Larrieu et al., 2002; Ritchie, Artero, & Touchon, 2001; Tuokko et al., 2003). Similar instability has been documented in clinical and community cohort studies, which tend to employ more detailed, and presumably more reliable, cognitive assessment protocols in the determination of the MCI classification. Rates of reversion in such studies have ranged from 7% to 44% over 1–2-year follow-up intervals (Bruscoli & Lovestone, 2004; de Jager and Budge, 2005; Devanand, Folz, Gorlyn, & Moeller, 1997; Loewenstein, Acevedo, Agron, & Duara, 2007).

The same factors that may account for variation in conversion rates may also contribute to variation in reversion rates in MCI (e.g., varying inclusion/exclusion criteria for the type of cognitive impairment and age of participants, use of clinical vs. community samples, differing methods of identifying cognitive impairment, varying reliability of cognitive assessment measures, sample sizes, statistical methods, length of longitudinal follow-up, etc.). An additional factor to consider is the potential role of practice effects (i.e., improved test scores at follow-up measurement occasions secondary to factors such as prior experience with test material and reduced anxiety in the assessment situation). An important and largely overlooked issue in MCI research that may be a significant contributor to classification instability may be the reliance on “single-session assessment” and “single-test impairment”. That is, although wide variation in the operational definitions of MCI exists across studies, in practice, the vast majority of the recent studies reviewed earlier specify an objective cognitive impairment inclusion criterion for MCI which is operationalized as “impaired performance on a single psychometric test at a single time point” (Busse et al., 2006; de Jager & Budge, 2005; Ganguli et al., 2004; Larrieu et al., 2002; Loewenstein et al., 2007; Ritchie et al., 2001). This practice, although common, is highly problematic. Large-scale normative data collection efforts have demonstrated that isolated “impaired” scores on neuropsychological measures are relatively common among “normal” samples. For example, within the carefully screened, neurologically normal sample employed by Heaton and colleagues (1991) to generate comprehensive normative data for the Halsted–Reitan neuropsychological battery, 90% of participants obtained at least one score in the “abnormal” range (i.e., T-score ≤39).

More recent data specific to older adults come from two studies by Brooks and colleagues (Brooks, Iverson, & White, 2007; Brooks, Iverson, Holdnack, & Feldman, 2008) who examined base rates of low memory scores among the neurologically normal older adults within the Neuropsychological Assessment Battery (NAB; Stern & White, 2003) normative sample (age 55–79, N = 742) as well as the Wechsler Memory Scale-Third Edition (WMS-III, Wechsler, 1997) normative sample (age 55–87, N = 550). On the NAB Memory Module, a battery of four memory tests that yields 10 subtest T-scores based on age-, gender-, and education-corrected norms, over half (55.5%) of the “normal” individuals had at least 1 of the 10 subtest scores >1 SD below their demographic group mean, 30.8% of individuals had one score >1.5 SD below the mean, and 16.4% had at least one score >2 SD below the mean (Brooks et al., 2007). Similarly, when considering just 4 of the 11 subtests on the WMS-III (yielding a total of eight norm-referenced scores), 70% of the “normal” individuals had at least one subtest score >1 SD below the mean for their demographic group and 21.6% had at least one of the eight scores >2 SD below the mean (Brooks et al., 2008). This phenomenon of highly prevalent “impaired” scores in healthy participants is not isolated to normed, multitest batteries. Palmer, Boone, Lesser, and Wohl (1998) documented a 73% rate of “impairment” (e.g., 1.3 SD below the normative mean; below the 9th percentile) in a healthy, neurologically normal, older adult sample (age 50–80; N = 132) across a selection of neuropsychological tests commonly employed in clinical practice.
Taken together, these findings suggest that current single-session assessment practices with single-test impairment inclusion criteria for MCI will likely lead to elevated rates of false positives that may, in turn, account for some of the heterogeneity of outcomes for individuals diagnosed as MCI. Two tactics may be employed to address these issues. First, multiple-session assessments may be employed to improve differentiation of those individuals with stable neuropsychological impairment or decline from those showing “accidental” (de Rotrou et al., 2005) poor performance on testing due to transient factors such as low mood or motivation, or fatigue. Second, multiple-test impairment inclusion criteria could be employed to improve differentiation of those individuals with robust neuropsychological impairment from those showing isolated low scores due to accidental poor performance. Of these two tactics, the former may be best suited to epidemiological and community-based longitudinal research studies where individual cognitive test batteries may be minimal, but data are collected at multiple time points. The latter may be better suited to clinical studies where emphasis may be placed on gathering more complete cognitive assessment data to make more reliable diagnoses of MCI. [A combination of these approaches may, of course, be applied. Such a strategy is implied in the current consensus classification criteria for MCI (i.e., Winblad et al., 2004), which specify that cognitive impairment may be identified in cases where there is a subjective cognitive complaint and objective cognitive impairment and/or where there is “evidence of objective decline over time.”]

To date, one existing study has employed a multiple-session assessment approach in a community sample. Collie, Maruff, and Currie (2002) performed repeat assessment of older adults semi-annually over a period of 1 year. They found that while roughly 20% of study participants met criteria for MCI (defined as performance 1.5 SD below the normative mean on a measure of delayed verbal recall) at any one testing session, only 13% met criteria at all three sessions. The latter, “Persistent MCI” group, captured those individuals with more consistent cognitive difficulty including those whose already low cognitive ability was on a declining trajectory. These authors suggested that individuals with Persistent MCI are likely at a greater risk of dementia than those showing transient poor performance during a single assessment. The current study was designed to replicate this work using slightly modified procedures and extend these findings by examining the longitudinal course of cognitive and functional outcomes for those identified as MCI across multiple assessments.

**Study Objective and Hypotheses**

Our primary study objective was to provide evidence for the external validity of a Multiple-Assessment (MA) MCI classification procedure. Using data from a prospective 5-year longitudinal study of cognitive change in community-dwelling older adults and procedures adapted from Collie and colleagues (2002), a psychometric algorithm was applied to a sample of community-dwelling older adults to identify the rate of cognitive impairment observed at a single-measurement occasion (i.e., “Single-Assessment [SA] MCI”) versus the rate of cognitive impairment observed at consecutive measurement occasions (i.e., “MA-MCI”). [We employ the term “SA-MCI” to refer to the general practice of identifying MCI based on single-session assessment as is the typical practice in the epidemiological and clinical research literature. This is inclusive of epidemiological methods for identifying MCI based on available psychometric test data at a single time point, as well as clinical methods for formally diagnosing MCI based on a single-session evaluation by a skilled practitioner.] Consistent with Collie and colleagues (2002), we expected that the prevalence of SA-MCI would exceed that of MA-MCI.

As prior studies have demonstrated that those with SA-MCI go on to develop dementia, a syndrome characterized by impairment in cognitive and functional status, at higher rates relative to controls, it was hypothesized that both the SA-MCI and the MA-MCI groups would show lower levels of performance and greater decline in both cognitive and functional status over 5 years relative to Controls. Further, and to the extent that the MA-MCI classification improves upon limitations in the SA-MCI classification, thereby more accurately capturing those at risk of dementia, it was hypothesized that the magnitude of difference in baseline level of performance and longitudinal trajectory of decline in cognitive and functional status between those classified as MA-MCI and Controls would exceed the magnitude of difference in baseline level of performance and longitudinal trajectory of decline in cognitive and functional status between those classified as SA-MCI and Controls.

**Materials and Methods**

**Participants**

Data for this study were drawn from Years 1 to 6 of a parent study, an ongoing prospective investigation of short-term fluctuations in performance and a long-term cognitive change in older adults. At baseline, participants were 304 Caucasian, healthy, community-dwelling older adults (208 women and 96 men), aged 64–92 (M = 74.0, SD = 6.0), recruited through local media advertisements, seeking individuals who were concerned about their cognitive functioning, but not diagnosed with an exclusionary neurological disorder. Exclusionary criteria included physician-diagnosed dementia or a Mini-Mental...
State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) score of < 24, a history of significant head injury (e.g., loss of consciousness > 5 min), other neurological or major medical illnesses (e.g., Parkinson’s disease, heart disease, cancer), severe sensory impairment (e.g., difficulty reading newspaper-size print, difficulty hearing a normal conversation), drug or alcohol abuse, current psychiatric diagnosis, psychotropic drug use, and lack of fluency in English.

Because the MA-MCI classification scheme under investigation in the current study required cognitive testing data from two serial annual assessments, 42 participants who completed Year 1, but not Year 2 cognitive assessment, were excluded from the final sample. The excluded participants did not differ from the final sample in terms of age, education, gender, health status (number of chronic conditions), or functional status (Activities of Daily Living, ADLs). However, the excluded participants had a slightly lower average MMSE (Folstein et al., 1975) score ($M = 28.4, SD = 1.7$), relative to the final sample ($M = 28.8, SD = 1.1$), $F(1, 303) = 4.392$, $p < 0.05$, $\eta^2 = 0.01$. The final sample comprised 262 adults (180 women and 82 men) aged 64–92 ($M = 73.8, SD = 5.8$), with a mean of 15.2 years of education ($SD = 3.1$).

Procedure

Potential participants were screened for inclusion and exclusion criteria by a telephone interview. Eligible participants completed two testing sessions (one group and one individual) scheduled over approximately 3 months. The group testing session was held in the laboratory, and the individual testing sessions were conducted in the participants’ homes. At the group session, participants provided demographic and health information and completed a series of cognitive benchmark measures. During the individual session, participants completed a series of individually administered cognitive and functional measures. Over the course of the parent study, these group and individual testing sessions were repeated annually at Years 1 through 4, and again at Year 6. During Year 5, the face-to-face testing sessions were replaced with a health update and cognitive screening completed via telephone.

Measures

Participant characteristics. Demographic information (age, gender, years of education, marital status, native language, ethnic background) was obtained from participants during the initial group testing session and verified annually during the individual sessions. Additional background information collected at baseline and included in the current study included self-reports of memory (Gilewski, Zelinski, & Schaie, 1990; items rated on a 7-point scale where 1 = much worse, 4 = same, 7 = much better), health (Hultsch, Hertzog, Dixon, & Small, 1998; items rated on a 5-point scale where 1 = very poor, 5 = very good), and depressive affect (Depressive Affect Subscale of Centre for Epidemiological studies Depression Scale, CES-D; Hertzog, Van Alstine, Usala, & Hultsch, 1990; Radloff, 1977; score range 7–28; higher scores indicate greater depressive affect).

Cognitive benchmark measures used for MCI classification. A series of cognitive benchmark measures were administered at annual group testing sessions. These measures were used to classify participants into MCI groups according to the algorithm outlined in the upcoming section titled “MCI Classification Procedures.”

The WAIS-R Digit Symbol Substitution task (Wechsler, 1981) was used to assess perceptual processing speed.

The Letter Series test (Thurstone, 1962) was used to assess inductive reasoning. A series of letters following a distinct pattern was presented and the participant was asked to decipher the pattern in the target string and provide the next letter in the sequence.

A word recall task (Hultsch, Hertzog, & Dixon, 1990) was used to assess episodic memory. One list of 30 English words selected from a total set of six lists was presented. The list contained six words from each of five taxonomic categories typed on a single page in unblocked order. The participant was given 2 min to study the list and 5 min to write the words, they could recall in any order.

The Controlled Associations test from the ETS kit of factor-references cognitive tests (Ekstrom, French, Harman, & Dermen, 1976) was used to assess verbal fluency. The participant was given 6 min to generate as many synonyms as possible in response to a set of target words.

A recognition vocabulary test was used to assess vocabulary. The test was composed by combining three 18-item tests from the ETS kit of factor-references cognitive tests (Ekstrom et al., 1976). The participant was given 15 min to complete a 54-item multiple-choice task.

Cognitive and functional outcome measures. The MMSE (Folstein et al., 1975) was used to survey global cognitive functioning in the domains of orientation, memory, attention, language, and visuoconstructive ability. The MMSE was administered at
study Years 1 through 4, and again at Year 6. During Year 5, the Telephone Interview for Cognitive Status (Brandt, Spencer, & Folstein, 1988) was used to generate MMSE equivalent scores. Higher scores on the MMSE indicate higher cognitive functioning.

The Trail Making Test, Part B (Trails B, Reitan & Wolfson, 1985) was used as a measure of executive functioning involving perceptual speed, sequencing, and mental flexibility. Time-to-task completion (seconds) was used as the measure of interest. Higher scores on this measure indicate poorer performance. This measure was administered at Years 1 through 4 of the project.

The WAIS-R Digit Symbol Substitution-Incidental Recall task (Coding Recall, Wechsler, 1981) was used annually to assess perceptual processing speed as outlined previously. Following the coding portion of the task, participants were presented with a sheet containing the nine symbols and asked to recall the number that had been paired with the symbol to provide an index of memory performance. This number of items drawn correctly (maximum = 9) was used as a measure of incidental visual recall.

To measure global functional status, participants were asked to rate their level of difficulty with nine basic and instrumental ADLs (i.e., walking across a room, bathing self, dressing self, getting up from a bed or chair, climbing stairs, walking several blocks, managing finances, performing household activities, and driving a car), on a scale from 1 to 3 (1 = no difficulty, 2 = some difficulty, 3 = a lot of difficulty; Rodgers & Miller, 1997). This measure was administered at Years 1 through 4, and again at Year 6. A total score was obtained by summing participants’ responses across the nine activities. Higher scores indicate greater difficulty with ADLs.

The Everyday Problems Test (EPT; Willis & Marsiske, 1993) is a paper and pencil test of everyday cognitive ability administered at Years 1 through 4, and Year 6 as a measure of everyday problem solving. This measure uses 21 printed stimuli designed to closely mimic items encountered in daily life (e.g., medication label, pay telephone instructions) and requires participants to solve problems pertaining to the stimuli. Test items cover major components of instrumental ADLs including medication use, meal preparation, telephone use, shopping, financial management, household management, and transportation. Higher scores indicate better performance.

**MCI classification procedures.** Classification of cognitive group status was determined based on performance on the five cognitive benchmark tasks (i.e., perceptual speed, reasoning, episodic memory, verbal fluency, and vocabulary.) Normative data were drawn from an independent sample of adults aged 65–94 recruited from the same population (Victoria Longitudinal Study; Dixon & de Frias, 2004). The use of local normative data derived for all tasks on the same population is preferred to the use of other published normative data given the close demographic match of the local sample to the current sample and the ability to make more accurate comparisons across tasks. Normative data for perceptual speed, reasoning, verbal fluency, and vocabulary tasks are based on data from 445 individuals (282 women and 163 men), and normative data for the episodic memory task are based on data from 194 individuals (125 women and 69 men).

Figure 1 is a schematic representation of the algorithm applied for classifying individuals according to the SA-MCI and the MA-MCI classification. The algorithm was based on current consensus criteria for identifying the presence of MCI (Winblad

![Fig. 1. Classification of Single- and Multiple-Assessment MCI.](image-url)
et al., 2004). Sample size limitations precluded further differentiation of MCI into commonly used subgroupings (e.g., amnestic vs. nonamnestic, single domain vs. multiple domain). As a starting point, participants were classified as having a cognitive impairment if they obtained scores $>1.5\ SD$ below the mean of their age- and education-matched peers on any one of the cognitive benchmark tasks. For the SA-MCI categorization, data from the Year 1 baseline cognitive assessment were considered. Individuals who demonstrated cognitive impairment at Year 1 were categorized as SA-MCI, and those who did not demonstrate impairment were categorized as Controls. For the MA-MCI categorization, data from the first two annual cognitive assessments were considered. Individuals who demonstrated cognitive impairment at each of the first two assessments (Years 1 and 2) were categorized as MA-MCI, those who demonstrated impairment at one, but not both time points, were categorized as Unstable, and those who did not demonstrate impairment at any point were categorized as Controls.

**Results**

**Preliminary Analyses**

Data regarding the prevalence of MCI are summarized in Table 1. As expected, base rates of impaired performance on any one cognitive test at any one time point in the current study (SA-MCI, $N = 77$, or 29% of sample) exceeded those of persistent impaired performance observed on consecutive assessment sessions (MA-MCI, $N = 75$, or 17% of sample). Of the 77 individuals who met the criteria for SA-MCI, 32 individuals (42% of the group) did not meet those for MA-MCI.

Exploratory comparison of cognitive status group differences in demographic and self-report characteristics revealed a trend for the Control groups to be younger, more educated, less depressed, and healthier relative to the MCI groups, though these differences were small and not consistently statistically significant. In general, the Unstable group tended to perform at a level intermediate to the Control and the MCI groups in terms of self-report characteristics, though these differences were generally not statistically significant. Globally, the study participants reported themselves to be a relatively well-educated ($M = 15.2$ years of education, $SD = 3.1$) and euthymic (CES-D $M = 8.0$, $SD = 2.2$) sample and tended to rate their own health ($M = 4.5$, $SD = 0.06$) as slightly better than their peers.

Subtle group differences in self-reported memory ability were observed. Self-rated memory change over the past year was rated as slightly more negative in the MCI groups (Year 1 and Year 2 Controls $M = 3.9$, $SD = 0.8$; SA-MCI $M = 3.5$, $SD = 1.2$; MA-MCI $M = 3.6$, $SD = 1.3$). Self-rated memory relative to peers did not differ significantly by group, with all groups rating their perceived memory ability as generally on par with their peers (Year 1 Controls $M = 4.3$, $SD = 0.9$; Year 2 Controls $M = 4.2$, $SD = 0.9$; SA-MCI $M = 4.0$, $SD = 1.2$; MA-MCI $M = 4.2$, $SD = 1.2$).

Preliminary analyses yielding descriptive statistics to characterize baseline performance in raw score units by cognitive group on each of the key outcome variables under consideration in the primary analyses are reported in Table 2.

**Primary Analyses**

The primary study objective addressed the external validity of the MA-MCI classification procedure. Hierarchical Linear Modeling (HLM) was used to examine cognitive status group differences in baseline level and longitudinal trajectory of each cognitive (MMSE, Trails B, and Coding Recall) and functional (ADL and EPT) outcome measures. HLM is a particularly appropriate means of analyzing longitudinal data as it can account for both within-person trajectories of performance over time, as well as between-person differences (e.g., in cognitive status group) in these individual trajectories (Raudenbush & Bryk, 2002). In the case of longitudinal data, this task is accomplished by modeling observed variance at two levels: within-person (Level 1) and between-person (Level 2). An added advantage of HLM models is that parameters are estimated using full information maximum likelihood (FIML) procedures. FIML offers a built-in means to preserve data from individuals who may have incomplete participation over the course of a longitudinal study by assigning increased weight to those with no missing data when calculating parameter estimates.

**Table 1. Prevalence of Single- versus Multiple-Assessment MCI**

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Unstable</th>
<th>MCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1: Single-Assessment MCI</td>
<td>185 (71%)</td>
<td>not defined</td>
<td>77 (29%)</td>
</tr>
<tr>
<td>Year 2: Multiple-Assessment MCI</td>
<td>162 (62%)</td>
<td>55 (21%)</td>
<td>45 (17%)</td>
</tr>
</tbody>
</table>

*Notes: MCI = Mild Cognitive Impairment. Values are $N$ with percentage of original sample ($N = 262$) in parentheses.*
The parameter intercept, or starting point. The parameter independent (Time in Study fit in order to estimate the sample average intercept and slope for the outcome variable of interest:

\[ Y_{ij} = \beta_0 + \beta_1 (\text{Time in Study})_{ij} + r_{ij} \]

where \( Y_{ij} \) is the value of the outcome variable of interest for person \( i \) at time \( j \), and \( \beta_0 \) represents the value of \( Y \) at an individual’s intercept, or starting point. The parameter \( \beta_1 \) represents the rate of change for an individual per single unit increase in the independent (Time in Study) variable. The parameter \( \gamma_0 \) is the average intercept for the sample, and \( \gamma_1 \) is the average slope for the sample. The parameter \( r_{ij} \) is an error term representing unexplained within-person variance, and the parameters \( u_0 \) and \( u_1 \) are error terms representing unexplained between-person variance.

Two separate “complex” models were then fit for each outcome variable. In the first complex model, a between-subjects cognitive status group variable was fit capturing individual membership in the Control versus the SA-MCI grouping:

\[ Y = \beta_0 + \beta_1 (\text{Time in Study}) + r \]

\[ \beta_0 = \gamma_{00} + u_0 \]

\[ \beta_1 = \gamma_{10} + u_1 \]

In the second complex model, a between-subjects cognitive status group variable was fit capturing individual membership in the Control versus the MA-MCI grouping:

\[ Y = \beta_0 + \beta_1 (\text{Time in Study}) + r \]

\[ \beta_0 = \gamma_{00} + \gamma_{01} (\text{Ctrl. vs. SA-MCI}) + u_0 \]

\[ \beta_1 = \gamma_{10} + \gamma_{11} (\text{Ctrl. vs. SA-MCI}) + u_1 \]

Parameter estimates for the two sets of complex model Level 2 equations were used to determine average baseline performance and trajectory of change in outcome measures by cognitive status group. To illustrate, parameter estimates for the MMSE models are shown graphically in Fig. 2. In the SA model (Fig. 2, left), the Control group had an average baseline MMSE score of 28.9 and an average rate of decline of −0.06 points per year. The SA-MCI group had an average baseline score of 28.6 and an average rate of decline of −0.21 points per year. In the MA model (Fig. 2, right), the Control group had an average baseline MMSE score of 29.1 and an average rate of decline of −0.06 points per year. The MA-MCI group had an average baseline score of 28.6 and an average rate of decline of −0.33 points per year. The unstable group had an average baseline score of 28.6 and an average rate of decline of −0.16 points per year.

Table 3 presents a summary of cognitive status group differences in baseline performance and rate of change over time as derived from HLM models outlined previously. The values are presented on a Z-score metric (i.e., \( M = 0, SD = 1 \); baseline mean and SD values for the Year 1 Control group used as a normative reference) to facilitate informal comparison of baseline and a rate of change differences across groups and across measures. For example, the difference between the SA-MCI group

<table>
<thead>
<tr>
<th>Cognitive Measures</th>
<th>Single-Assessment</th>
<th>Multiple-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>SA-MCI</td>
</tr>
<tr>
<td>MMSE</td>
<td>28.9 (1.1)</td>
<td>28.6 (1.3)</td>
</tr>
<tr>
<td>Trails B</td>
<td>85.3 (29.9)</td>
<td>98.7 (36.7)</td>
</tr>
<tr>
<td>Coding Recall</td>
<td>6.4 (2.2)</td>
<td>4.9 (2.6)</td>
</tr>
<tr>
<td>Functional Measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activities of Daily Living</td>
<td>10.3 (2.1)</td>
<td>11.0 (2.8)</td>
</tr>
<tr>
<td>Everyday Problems</td>
<td>32.4 (5.0)</td>
<td>28.1 (7.1)</td>
</tr>
</tbody>
</table>

Notes: SA-MCI = Single-Assessment Mild Cognitive Impairment; MA-MCI = Multiple-Assessment Mild Cognitive Impairment; MMSE = Mini-Mental State Examination. Values are means with standard deviations in parentheses. For all measures except Trails B and Activities of Daily Living, higher scores indicate better performance. N = 262.

To test hypotheses associated with the external validity of the MA-MCI classification, a series of HLM models were run for each outcome measure. In each case, an initial “simple” model with a single within-person time predictor (Time in Study) was fit in order to estimate the sample average intercept and slope for the outcome variable of interest:
and Controls on the MMSE at baseline was equal to 0.26 $SD$ (relative to the Control group baseline performance). The difference between the SA-MCI group and Controls in terms of a 5-year rate of change, on the other hand, was equal to 0.97 $SD$.

The first set of hypotheses in this study objective posited that both the SA-MCI and the MA-MCI groups would show lower levels of performance and greater decline in both cognitive and functional status over 5 years relative to Controls. As shown in Table 3, a similar pattern was observed across most measures. That is, baseline performance was significantly lower in both the SA-MCI and the MA-MCI groups compared with Controls. (An isolated exception was a failure to find a significant difference

Table 3. Standardized absolute baseline and 5-year change differences for Controls versus Single-Assessment and Controls versus Multiple-Assessment MCI groups

<table>
<thead>
<tr>
<th>Cognitive Measures</th>
<th>Single-Assessment MCI</th>
<th>Multiple-Assessment MCI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Five-year change</td>
</tr>
<tr>
<td>MMSE</td>
<td>0.26</td>
<td>0.97**</td>
</tr>
<tr>
<td>Trails B</td>
<td>0.57**</td>
<td>0.82**</td>
</tr>
<tr>
<td>Coding Recall</td>
<td>0.70**</td>
<td>0.03</td>
</tr>
<tr>
<td>Functional Measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activities of Daily Living</td>
<td>0.32*</td>
<td>0.08</td>
</tr>
<tr>
<td>Everyday Problems</td>
<td>0.89**</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Notes: MCI = Mild Cognitive Impairment; MMSE = Mini-Mental State Examination. In all cases where results were statistically significant, the MCI groups performed more poorly at baseline and/or had a steeper trajectory of decline over time relative to the controls.

*To facilitate informal comparison across measures, values are presented on a common $Z$-score metric, using Year 1 Control group baseline performance as a normative reference. For example, the difference between the SA-MCI group and Controls on the MMSE at baseline was equal to 0.26 $SD$, while the difference between the SA-MCI group and Controls in terms of 5-year rate of change was equal to 0.97 $SD$.

Five-year rate of change for Trails B data is extrapolated from Years 1 to 4 change data.

$p < .05$.

**$p < .01$. 

Fig. 2. Group mean performance (HLM Level 2 model parameter estimates) on the MMSE by year for the Single-Assessment MCI classification (left) and the Multiple-Assessment MCI classification (right).
on the MMSE between the SA-MCI and Control groups.) In terms of rates of longitudinal decline, statistically significant findings were fewer. The expected finding that the SA-MCI and the MA-MCI groups would decline more rapidly than Controls was only seen on the MMSE and the Trails B measure. It is noted that, at the sample level, a significant negative overall trajectory of change was observed for all measures except the Coding Recall measure, which had an overall trajectory of change that did not differ significantly from zero.

The second set of hypotheses in this study posited that the magnitude of discrepancies in baseline performance and the rate of change over time between MA-MCI versus Controls would exceed discrepancies between SA-MCI versus Controls. Comparison of overall model fit parameters between the two complex models demonstrated lower Akaike’s Information Criteria values (representing better model fit) for the MA-MCI model relative to the SA-MCI model. Statistical comparison of the relative magnitude of the Level 2 parameter estimates was not possible as the two complex models are non-nested. However, descriptive comparison of the Level 2 parameter estimates (shown in Table 3) supported the outlined hypothesis. That is, the difference in baseline performance between the MA-MCI group and Controls was larger than the difference in baseline performance between the SA-MCI group and Controls on all of the cognitive and functional measures. Similarly, the difference in rate of change between the MA-MCI group and Controls was larger than that in the rate of change between the SA-MCI group and Controls.

Discussion

The present study was designed to address key limitations in current MCI classification procedures which tend to rely on single occasion assessment (“SA-MCI”) by evaluating an alternate operational definition of MCI requiring evidence of persistent cognitive impairment over multiple-testing sessions (“MA-MCI”). Broadly, study findings demonstrate that an MA-MCI group showed a poorer trajectory of cognitive, and to a lesser extent, functional decline over a 5-year follow-up interval.

It was hypothesized that both the SA-MCI and the MA-MCI groups would show poorer cognitive and functional performance than Controls at baseline and in terms of longitudinal change. This hypothesis was almost universally supported in terms of cognitive status group differences in baseline performance. In terms of group differences in the rate of change over time, significant findings were more frequent among the cognitive, when compared with the functional, measures. Both the SA-MCI and the MA-MCI groups showed a greater rate of decline on the MMSE and the Trails B measure compared with Controls.

To the extent that the MA-MCI classification improves upon limitations in SA-MCI classification, thereby more accurately capturing those at risk of dementia, it was further hypothesized that the magnitude of difference in the baseline level of performance and longitudinal trajectory of decline in cognitive and functional status between those classified as MA-MCI and Controls would exceed the magnitude of difference in the baseline level of performance and the longitudinal trajectory of decline in cognitive and functional status between those classified as SA-MCI and Controls. Although formal statistical evaluation of the relative magnitude of these group differences was not possible in the present design, descriptive comparison of the relevant parameter estimates supported this hypothesis.

The prevalence of SA-MCI (29% of sample) exceeded that of the MA-MCI (17% of sample). The proportion of individuals who met criteria for MCI at baseline (i.e., met criteria for SA-MCI) but did not meet criteria for MA-MCI was 41%; a value roughly in keeping with the 35% found over a 6-month follow-up interval by Collie and colleagues (2002) and the 49% found over a 1-week follow-up interval by Duff and colleagues (2008), and toward the high end of other previously reported rates of reversion in epidemiological and community samples, which range from 15% to 44%. Prior findings have shown that rates of instability of classification appear to be highly susceptible to the specific inclusion/exclusion criteria for MCI under consideration.

The high rates of reversion in the present study are unexpected given the commonly held assumption of generally declining (or at best, stable) cognitive abilities in aging. One key candidate contributor to this finding may be retest or practice effects. Recently, a number of authors have employed a range of novel statistical and methodological approaches targeted toward disassociating any presumably positive effect of prior testing from the presumably negative effects of age-related maturation in longitudinal studies of cognitive aging (e.g., Ferrer, Salthouse, Stewart, & Schwartz, 2004; Rabbitt, Diggle, Smith, Holland, & Innes, 2001; Ronnlund, Nyberg, Backman, & Nilsson, 2005; Thorvaldsson, Hofer, Berg, & Johansson, 2006). This work has provided evidence for a hypothesis that retest effects may account for a substantial portion of the historical discrepancy between the larger apparent age-related decline in cognitive ability evident in cross-sectional when compared with longitudinal studies. As Salthouse (2009) has recently argued, although it is commonly accepted that cohort effects (i.e., differences in performance as a function of birth cohort, with a trend for better performance in more recently-born cohorts) may lead to an overestimation of age-related decline in cross-sectional studies, recent advances in studying retest effects, as well as
converging evidence from animal and neurobiological studies, suggest that retest effects may lead to an underestimation of age-related decline in longitudinal studies.

This line of research has clear implications for longitudinal studies of MCI. As Ferrer and colleagues (2004) highlight, benefit from retest may be an individual characteristic. Indeed, a failure to achieve normative benefit from prior exposure to testing material may indicate an elevated risk of impending cognitive decline, despite an apparent stability of cognitive test scores over time. One prior study has provided some preliminary evidence in support of this hypothesis, illustrating that a previously identified MCI group, compared with healthy controls, showed significantly reduced retest benefit across four serial computerized testing batteries repeated within a single day (Darby, Maruff, Collie, & McStephen, 2002). However, Duff and colleagues (2008) documented preferential retest benefit in an amnestic MCI group, compared with Controls, on delayed recall measures administered at a 1-week interval. Longitudinal investigation of these issues warrants future study.

The substantial rate of reversion within the current and prior MCI studies supports the need for continued efforts to refine diagnostic criteria for MCI. The present study modification (i.e., requiring evidence of impairment across multiple assessments for MCI inclusion) represents one approach. Study findings support the feasibility, as well as provide some evidence for the efficacy of this approach for enhancing identification of those at greater risk of decline. Other possible modifications for future study include disregarding baseline testing to minimize the classification of those individuals who score poorly on initial assessment solely due to transient factors such as test anxiety associated with the novel assessment scenario. Another option, commonly employed in clinical assessment, is to compare domain-specific cognitive test scores to estimates of premorbid intellectual functioning.

Limitations of the present study include the degree to which the study classification procedures represent a less detailed assessment than would be typical of a full clinical diagnostic evaluation (e.g., the use of one test per cognitive domain, the use of an immediate, as opposed to delayed, memory measure to identify memory impairment). However, prior studies employing detailed assessment comparable to typical clinical evaluations have documented significant rates of reversion from MCI to Unimpaired status across repeat assessment similar to reversion rates reported in the epidemiological literature and the present study (de Jager & Budge, 2005; Devanand et al., 1997; Loewenstein et al., 2007), suggesting that there may be a considerable degree of overlap in factors that underlie classification instability in both the clinical and the research setting. Further limitations of the present study include the nature of the study sample (i.e., healthy, high-functioning volunteers), the application of specific inclusion/exclusion criteria for MCI classification (future research could investigate others of interest, such as amnestic vs. nonamnestic MCI; single- vs. multiple-domain MCI), and the availability of relevant data. The latter issue includes limitations in the available outcome measures, such as the use of Digit Symbol Coding Recall, a test not commonly used in the MCI literature, as a sole memory measure, though there is evidence that this measure discriminates MCI from Controls (Troyer et al., 2008). Further, limitations in the available data include the absence of data in relevant cognitive areas such as episodic verbal delayed recall, confrontation naming and visualspatial processing, as well as the absence of functional status information derived from informant report. These areas represent promising avenues for future study.

In sum, results of the present study demonstrate the feasibility and potential utility of an alternate operational definition of MCI requiring evidence of persistent cognitive impairment over multiple-testing sessions (MA-MCI). Results demonstrated that a group of individuals classified as MA-MCI showed a poorer trajectory of cognitive, and to a lesser extent, functional decline over a 5-year follow-up interval. Globally, results support the utility of the MA-MCI concept as an avenue to enhance detection of dementia risk and suggest a number of avenues for future study, including replication and extension of the present findings using alternate outcome measures, as well as more formally accounting for instability of classification related to retest effects in both the clinical setting and in future MCI studies.

Funding

This work was supported by grants from the Canadian Institutes for Health Research to DFH, The Alzheimer Society of Canada to ES, and the Natural Sciences and Engineering Research Council to DFH and ES. SV is supported by a Postdoctoral Fellowship from the Alzheimer Society of Canada. SWSM is supported by a Career Investigator Scholar Award from the Michael Smith Foundation for Health Research.

Conflict of Interest

None declared.
ES passed away in June, 2009, prior to the completion of this article. We thank Dr Roger Dixon for the use of data from the Victoria Longitudinal Study to establish norms for the classification of cognitive status groups of the present sample.

References


