The Relationship Between Short-Term Intraindividual Variability and Longitudinal Intraindividual Cognitive Change in Older Adulthood: Covariation and Prediction of Change

by

Allison Anne Marie Bielak
B. A., University of Winnipeg, 2002
M. Sc., University of Victoria, 2004

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

DOCTOR OF PHILOSOPHY

in the Department of Psychology

The Relationship Between Short-Term Intraindividual Variability and Longitudinal Intraindividual Cognitive Change in Older Adulthood: Covariation and Prediction of Change

by

Allison Anne Marie Bielak
B. A., University of Winnipeg, 2002
M. Sc., University of Victoria, 2004

Supervisory Committee

Dr. David F. Hultsch, Supervisor
(Department of Psychology)

Dr. Esther Strauss, Departmental Member
(Department of Psychology)

Dr. Michael A. Hunter, Departmental Member
(Department of Psychology)

Dr. Ryan Rhodes, Outside Member
(School of Exercise Science, Physical and Health Education)

Supervisory Committee

Dr. David F. Hultsch, Supervisor
(Department of Psychology)

Dr. Esther Strauss, Departmental Member
(Department of Psychology)

Dr. Michael A. Hunter, Departmental Member
(Department of Psychology)

Dr. Ryan Rhodes, Outside Member
(School of Exercise Science, Physical and Health Education)

Abstract

This dissertation presents two studies of intraindividual variability in a longitudinal context to further explore the relationship between short-term intraindividual variability and longer-term cognitive change in older adults. A sample of 304 community-dwelling older adults initially aged 64-92 years completed between 1 to 6 waves of annual testing over a 5-year period. Participants completed an extensive battery of accuracy- and latency-based tests covering a wide range of cognitive complexity. The first study addressed the longitudinal nature of intraindividual variability over 3 years. Group-based increases in inconsistency were limited to the latter half of older adulthood (i.e., 75 years and older), but there were significant individual differences across the entire sample. The covariation relationships between change in cognition and change in inconsistency were significant across the one-year interval, and found to remain stable across both time and older age. For each additional unit increase in intraindividual variability, participants' cognitive performance correspondingly declined. The strength of the coupling relationship however was stronger for fluid cognitive domains such as memory,

reasoning, and processing speed, and variability based on moderately and highly complex tasks provided the strongest prediction. Building on these results suggesting that intraindividual variability is highly sensitive to even subtle changes in cognitive ability, the second study addressed the capacity of intraindividual variability to predict cognitive ability and other meaningful change outcomes 5 years later. Inconsistency at Wave 1 was particularly sensitive to changes reflecting the early behavioural characteristics of dementia, including episodic memory ability, cognitive status, and attrition. In each case, greater inconsistency at baseline was associated with a greater likelihood of being in a maladaptive group 5 years later. Mean rate of responding was a comparable predictor of change in most instances, but differences emerged according to the complexity of their derived tasks. Variability based on moderate to high cognitively challenging tasks appeared to be the most sensitive to longitudinal changes in cognitive ability, and was uniquely predictive of the rate of attrition compared to neuropsychological tasks. These findings are promising of the potential utility and applicability of intraindividual variability in understanding and predicting intraindividual cognitive change in older adulthood.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Acknowledgements

Although it feels surreal to be at the end of my doctoral training, I know a lot of people saw this moment clearly all along. My first and foremost thanks go to my supervisor, Dr. David Hultsch, who expertly guided my progress and patiently helped me to develop my scholastic talents into what they are today. Thanks to his mentorship and advice, I know my future academic career will be a promising one. My sincere gratitude also goes to Dr. Esther Strauss, Dr. Michael Hunter, and Dr. Roger Dixon, who have each played a special supportive role in helping me develop and learn over the years. As a student who was initially unsuccessful in securing external funding, I am extremely grateful for the financial support offered to me by the Michael Smith Foundation for Health Research, the BC Medical Services Foundation, the Canadian Institutes of Health Research, and various local University donors. The completion of the statistical analyses herein and particularly my understanding of them would not have been possible without the guidance of Dr. Stuart MacDonald, who was extremely generous with his time, patience, and inspiring words. I am also appreciative to the staff at the Victoria Longitudinal Study, Arlene from Project MIND, and the committed participants of both longitudinal studies. I must acknowledge those around me day in and day out, including my wonderful office and lab mates over the years, and friends who were always willing to discuss ideas and share experiences and advice. Finally, to those closest to me, I cannot thank you enough. My parents were my long-distance cheering squad, and my boyfriend Mark's unwavering belief in my abilities and constant encouragement kept me on time and pushing forward. Thank you for always believing in me.

Introduction

Why is Intraindividual Variability Important?

Researchers investigating cognitive functioning in adulthood tend to examine adults' mean level of performance on various cognitive tasks. Such measurement allows easy comparison across individuals of different age groups, or calculation of within-person longitudinal change over time. However, this type of measurement consequently assumes that individuals' performance is also stable in the short-term, or from one moment to the next. In fact, there are theoretical arguments against this assumption. The life-span development perspective states that an individual is in constant fluctuation as a result of living in a dynamic environment (Nesselroade, 1991), and that stability is only a temporary break from the ongoing variations in functioning (Nesselroade & Featherman, 1997). Therefore, the study of developmental change has to include and expect both variability and stability in development (Nesselroade & Featherman, 1997). These arguments suggest that research investigating the mean level of performance in older adulthood has only been considering part of the story.

Nesselroade (1991) clarified these two developmental distinctions further by distinguishing between two types of change: *Intraindividual change* is defined as changes that occur relatively slowly and over a relatively long period of time (e.g., year-to-year), and result in long-lasting changes in an individual (e.g., developmental change, learning of skills). On the other hand, *intraindividual variability* is defined as changes that occur relatively quickly and over relatively short time frames (e.g., moment-to-moment; week-to-week), and are temporary or reversible. Examples include shifts in mood, and fluctuations in physical or cognitive performance. A further distinction between the two types of change is statistical, in that intraindividual changes represent systematic change,

whereas intraindividual variability is unsystematic. For example, on a reaction time task a person's intraindividual change would be represented by their beta weight, or statistical change over time, but their intraindividual variability would be represented as the variation in performance around their own personal regression line. However, a dynamic relationship between the two change types likely exists, where short-term variations play a role in the occurrence of long-term developmental change (Nesselroade, 1991). Therefore, changes in the distributions of intraindividual variability may be indicators of impending longitudinal intraindividual change (Hultsch, Strauss, Hunter, & MacDonald, 2008). This dissertation evaluated these hypotheses by examining whether intraindividual variability changes with and is predictive of later longitudinal outcomes.

Various Definitions of Intraindividual Variability

Although intraindividual variability is defined as short-term but reversible change, further distinctions of the concept can be made along various dimensions. First, the time frame against which short-term change occurs is particularly important. For example, it is possible for transient change to exist across moments, trials, days, weeks, or even months. This diversity is evident in the research literature, as some studies have focused on particularly rapid change across trials of a reaction time (RT) task (e.g., Hultsch, MacDonald, & Dixon, 2002), some have focused on day-to-day fluctuations (e.g., Sliwinski, Smyth, Hofer, & Stawski, 2006), and others have investigated performance fluctuations across weeks (e.g., Eizenman, Nesselroade, Featherman, & Rowe, 1997). A handful of studies have even compared variability measurements based on different time frames. For example, Hultsch, MacDonald, Hunter, Levy-Bencheton, and Strauss (2000), Nesselroade and Salthouse (2004), and Rabbitt, Osman, Moore, and Stollery (2001), all found positive correlations between trial-based variability and session-based variability;

individuals who were more variable from trial-to-trial were also more variable from session-to-session. However, Rabbitt and colleagues noted that the trial-based intraindividual variability did not completely account for the session-based intraindividual variability, indicating that the underlying causes of variability may be slightly different depending on their derived time frame.

This leads into the second important distinction of intraindividual variability; there are varying potential sources of influence or causes. For example, weekly variability is more likely to be influenced by external factors like stress or fatigue than variability occurring across the trials of a task. Rather, the causes of rapid variability across trials are more likely to be endogenous, or derived from causes within the individual, such as the connectivity of neural networks or the efficiency of neurotransmitters. In fact, Martin and Hofer (2004) described a range of possible causes for variability based on its time frame, from attentional lapses for moment-based measures, fatigue, motivation, and order effects for session comparisons, and environmental causes, physical health, or practice effects for daily or weekly fluctuations. Therefore, it is important that the type of intraindividual variability being used reflects the source of influence one hopes to investigate.

A third distinction involves the scope of intraindividual variability (Li, Huxhold, & Schmiedek, 2004). Is the variability measured as fluctuations in performance on a single task (univariate), or is the variability defined as changes in the organization of abilities while performing a number of tasks (multivariate)? Li and colleagues noted that variations on a single task most likely represent endogenous mechanisms such as fluctuations in processing, but analyzing intraindividual variability in a dual-task

paradigm most likely represents switches in resource allocation, compensation, and organization.

Finally, intraindividual variability can be viewed as either adaptive or maladaptive, in that it is associated with positive or negative outcomes. Siegler (1994) described intraindividual variability in childhood as critical in promoting cognitive change, suggesting that the fluctuations act as evidence of trying different strategies for completing a task. He found that the trial immediately preceding the discovery of a new strategy and the trial on which the discovery was made were more variable. Therefore, in situations of learning, intraindividual variability may be adaptive. Similarly, Allaire and Marsiske (2005) found intraindividual variability was positively correlated with cognitive performance in older adults. According to Li, Huxhold et al. (2004), the adaptability of greater intraindividual variability appears to depend on a few key factors. Variability in responding is adaptive if the optimal way to perform the task is unknown, thus requiring one to experiment with various strategies, and there is consequently substantial room for growth and improvement on the task. This situation is more likely to occur on tasks that focus on accuracy rather than speed. However, there are limitations to the adaptive nature of intraindividual variability; once the level of learning asymptotes and peak performance is reached, continued intraindividual variability in responding is deemed to be maladaptive. Alternatively, intraindividual variability in responding on tasks that have little room for improvement (i.e., performance is already near ceiling), and where the current method of completing the task is appropriate, is considered maladaptive. This situation typically describes most perceptual speed or RT tasks. Accordingly, a large body of research investigating intraindividual variability on RT tasks has found greater variability is associated with poorer levels of performance on other cognitive tasks (e.g.,

Hultsch et al., 2002, Rabbitt et al., 2001). Overall, the type of task employed may play a role in the type of variability observed.

For the present studies, we were interested in the hypothesis that intraindividual variability reflects brain-based fluctuations, and will consequently be defining intraindividual variability as moment-to-moment changes or fluctuations in performance across trials. We also used a univariate scope, and focused on the theory that intraindividual variability is maladaptive and thus indicative of brain-based dysfunction. This type of variability has also been termed inconsistency (Hultsch et al., 2000) and both terms will be used interchangeably throughout this dissertation.

Reliability of Intraindividual Variability

Despite the varying definitions of intraindividual variability, and arguments about its essential role in the developmental process, what information can be gained by studying inconsistency in performance? Fluctuations around a person's mean performance are usually deemed to only represent error in responding because inconsistency appears to be random. But are there any logical patterns within intraindividual variability in performance?

In fact, even within a concept representing change and development, research has revealed that there is considerable stability in inconsistency. That is, the amount of fluctuation in cognitive performance has been shown to be a relatively stable characteristic of an individual. This evidence can be divided into four domains: the stability of intraindividual variability, its relationship to other stable characteristics, variability within stable characteristics, and the neurological correlates of intraindividual variability.

Stability of Intraindividual Variability

A number of studies have compared the amount of intraindividual variability present on different occasions of the same task, and across similar but slightly different tasks. First, if inconsistency is a stable trait of an individual, the amount of fluctuation in performance at one point in time should be positively correlated with the amount of fluctuation in performance at another point in time, and also with the amount of fluctuation seen on similar tasks. On the other hand, if inconsistency in cognitive performance is completely random, the measurements should be unrelated. In a comparison including healthy, arthritic, and demented older participants, Hultsch et al. (2000) found that individuals who were more variable across tasks on one occasion, also tended to be more variable across tasks on other occasions, and those who were more variable trial-to-trial on one RT task, were also more variable trial-to-trial on other RT tasks. Similar findings were found by Fuentes, Hunter, Strauss, and Hultsch (2001), Hultsch and colleagues (2002), Rabbitt and colleagues (2001), and Allaire and Marsiske (2005). Further, positive correlations have even been found for comparisons which computed separate reliability measures for odd and even trials (e.g., Jensen, 1992) or split the occasions in half and calculated two reliability statistics (Eizenman et al., 1997).

Next, if inconsistency in cognitive performance is characteristic to an individual, the amount of variability on one cognitive task should be positively correlated with the amount of variability on another cognitive task. The positive link would be expected if intraindividual variability were the cause of relatively stable endogenous mechanisms such as neurological dysfunction that were present during the performance of each task, rather than relatively transient exogenous influences such as fatigue or stress that may be present during one task and not the other. Research results have been consistent with this interpretation, and found individuals who showed more fluctuation in performance on one

RT task tended to show more fluctuation on other RT tasks (Hultsch et al., 2000; 2002; Fuentes et al., 2001). Relatedly, Hultsch and colleagues (2002) found a significant positive relationship between two types of intraindividual variability: those with greater dispersion or intraindividual variability in performance *across tasks* tended to also show greater inconsistency *across trials* on all tasks.

Finally, regardless of the stability of intraindividual variability in cognitive performance, is it of sufficient magnitude to warrant interest and study, particularly relative to other types of variability? Nesselroade and Salthouse (2004) argued that if intraindividual variability is small relative to interindividual differences, the phenomenon may only be of theoretical interest. On the other hand, if intraindividual variability is of substantial size, and not simply error variance, its presence demands an explanation. Nesselroade and Salthouse (2004) investigated this issue by comparing three types of variability in adults' performance on perceptual-motor tasks: a) between-person variability; b) within-person variability across trials of the task; and c) within-person variability across three occasions of the task. They found that the size of intraindividual variability was about one half as large as that of interindividual variability, a conclusion consistent with later findings on accuracy tasks (Salthouse, Nesselroade, & Berish, 2006), and memory and sensorimotor domains (Li, Aggen, Nesselroade, & Baltes, 2001). Further, by manipulating the regression equation, Nesselroade and Salthouse (2004) inferred that the average within-person variability for that particular task was approximately equal to the variation expected to occur across a 26-year period of normal aging. Ram, Rabbitt, Stollery, and Nesselroade (2005) also found that there were substantial individual differences in intraindividual variability in intraindividual *change* across 36 occasions of a cognitive task. Individuals varied in how consistent they were in

responding at the first occasion of the task, their change in intraindividual variability across the occasions (e.g., some individuals improved more quickly than others), and the final level of inconsistency in performing the RT task, further demonstrating the individual nature of intraindividual variability. Clearly, inconsistency is of significant magnitude and stability to warrant interest in development.

Intraindividual Variability and Various Stable Characteristics

A critical test of intraindividual variability is whether it is related to stable individual characteristics, which could only occur if there was some constancy to the pattern of intraindividual variability displayed. Further, the relationship would need to occur in a way consistent with underlying hypotheses and theory. For example, if inconsistency is believed to be an indicator of maladaptive development, the relationship between adaptive developmental outcomes such as high cognitive ability and inconsistency should be negative (i.e., higher test score, lower inconsistency in performance). In fact, there have been numerous studies demonstrating significant relationships between intraindividual variability and an impressive range of ability and theoretical domains.

The first and most pronounced correlate of inconsistency in performance is with intelligence and cognitive ability. Early studies noted negative correlations of intraindividual variability with general intelligence, in that low aptitude individuals were excessively variable from one RT trial to the next compared to brighter individuals (Jensen, 1982; 1992). The expected relationship with IQ was also evident for both within-session and across-session variability (Rabbitt et al., 2001). In fact, a later growth curve analysis on the same data showed that those with higher intelligence reached lower asymptotic levels of inconsistency with practice on the RT task (i.e., they reached higher

levels of proficiency as evidenced by their greater consistency in responding), and regardless of practice effects, did not vary as much week-to-week compared to those with lower scores (Ram et al., 2005). Similar findings have been found in relation to cognitive tasks covering a range of cognitive domains. Hultsch and colleagues (2002) found more inconsistent performance on four different RT tasks was associated with lower scores on tasks assessing perceptual speed, working memory, episodic memory, and crystallized abilities. Greater inconsistency in RT performance was even predictive of poorer everyday problem solving abilities, such as the ability to decipher nutritional information or transportation cost (Burton, Strauss, Hunter, & Hultsch, in press).

The negative relationship with inconsistency in cognitive performance has even extended beyond performance indicators such as cognitive ability to include adaptive lifestyle behaviours. Bielak, Hughes, Small, and Dixon (2007) found older adults who had a higher frequency of participating in leisure activities such as completing crosswords, using the computer, or playing a musical instrument demonstrated less intraindividual variability on four RT tasks. Therefore, the cognitive benefits of maintaining an active lifestyle, or the "use it or lose it" hypothesis of cognitive aging were even evident in older individual's fluctuations in their cognitive performance. Further, inconsistency at baseline testing was predictive of later attrition in a 6-year longitudinal study of aging. MacDonald, Hultsch, and Dixon (2003) found significant differences between returnees and dropouts on simple, lexical decision, and semantic decision RT tasks. Individuals who discontinued their participation at Wave 1 were more inconsistent in responding than those who returned for the next two waves of testing. MacDonald and colleagues noted that it is well-known that attrition is not random, but tends to reflect underlying influences such as disease and cognitive impairment.

Therefore, intraindividual variability may be a sensitive indicator of these impending problems. Clearly, intraindividual variability is a stable characteristic of an individual, but higher amounts of inconsistency are maladaptive, as evidenced by poorer cognition, poorer health behaviours, and increased risk of drop-out, which can be an indication of a variety of underlying health problems (e.g., deteriorating health, abnormal aging and disease, or impending death). In fact, a recent study by MacDonald, Hultsch, and Dixon (in press) verified these conclusions by finding that inconsistency significantly increased per additional year closer to death.

Intraindividual Variability in Other Domains

Research has shown that significant amounts of intraindividual variability also exist in other functioning domains presumed to stable. For example, Li and colleagues (2001) asked older adults to complete three walking tasks biweekly for 7 months (i.e., turn 360 degrees, walk 10 feet at a normal pace, and walk 10 feet at a fast pace). They found significant within-person fluctuation in older adults' sensorimotor performance; the magnitude was approximately half the magnitude of interindividual differences. Furthermore, sensorimotor inconsistency was negatively correlated with level of performance on the walking tasks (i.e., longer time, more steps to turn, and slower walking pace), and text and spatial memory. Sensorimotor inconsistency was of similar magnitude, and reflected the same pattern of negative relationships that have been shown with cognitive inconsistency (e.g., Nesselroade & Salthouse, 2004; Hultsch et al., 2000), indicating inconsistency may be an stable individual attribute that influences performance regardless of ability domain.

Further evidence for this conclusion has been found in the domain of self-reported beliefs. A seminal study by Eizenman et al. (1997) assessed the week-to-week

fluctuation in older adults' perceived competence and locus of control. They found substantial variability in the weekly measurements, but the amount of variability was a stable individual difference as substantiated by high correlations among the beliefs indices when the 25 occasions were split in half. Moreover, Eizenman and colleagues found high perceived control did not predict mortality 5.5 years later, but the consistency in one's control beliefs did! Those who were more variable on the control measures had an increased likelihood of dying 5.5 years later. Therefore, the trait of inconsistency appears to be indicative of a vital underlying endogenous mechanism. Bielak, Hultsch, Levy-Ajzenkopf et al. (2007) also found significant intraindividual variability in older adults' reported general perceived control beliefs (i.e., locus of control) and memory-specific control beliefs (i.e., ways of improving their memory) across 10 bimonthly occasions. Despite the fact that participants completed these measures both before and after completing a battery of cognitive tasks, there was little evidence that these short-term changes were driven by changes in their actual performance.

Finally, some researchers have investigated the link between the variability found in one domain of performance (i.e., cognitive) with that found in another domain (i.e., physical, self-perceived affect/beliefs). Strauss, MacDonald, Hunter, Moll, and Hultsch (2002) hypothesized that if intraindividual variability was caused by endogenous mechanisms, the fluctuations in performing a cognitive task would be related to fluctuations in performing other domains. Specifically, inconsistency in physical performance (e.g., blood pressure, respiratory function) would correspond to inconsistency in cognitive performance if the same mechanism (i.e., brain-based) was the underlying cause of both expressions of variability. Therefore, just like cognitive inconsistency, physical inconsistency would consequently show links to cognitive level of

ability. On the other hand, the instability in an individual's perceptions, beliefs, and affect would show a weaker relationship with the index of cognitive inconsistency, because such variability was likely influenced by exogenous mechanisms such as stress and mood. Three different groups of older adults (i.e., those with mild dementia, arthritis, and healthy controls) completed a variety of physical (e.g., balance/gait, fine motor dexterity), affective/belief-based (e.g., positive and negative affect, perceived competence and control), and RT tasks over four weekly sessions. There was evidence of the expected cross-domain links between inconsistency in physical functioning and inconsistency in cognitive performance for all groups on simple cognitive tasks, but only for those with dementia on more challenging tasks. Further, increased inconsistency in non-cognitive domains was generally associated with poorer cognitive function. For example, greater intraindividual variability in diastolic blood pressure and non-dominant finger tapping was associated with poorer memory performance. However, while physical inconsistency uniquely predicted level of cognitive performance, variability in affect and perceptions did not. Overall, the positive association between cognitive and physical inconsistency, and the dissociation between the physical and affect/beliefs domains, added significant support to inconsistency's role as a stable, endogenous characteristic of an individual.

Neurological Correlates of Intraindividual Variability

Perhaps the most striking evidence that intraindividual variability represents a stable individual characteristic is its relationships with neurological conditions. Generally speaking, individuals who have a neurological condition or disease tend to show greater amounts of inconsistency in their behavioural responding compared to healthy individuals. References to this link were made as early as 1926, as Henry Head described

variability in behaviour as "one of the most striking results produced by a lesion of the cerebral cortex." This relationship has been explored and consistently demonstrated in a variety of clinical pathologies including blunt head injuries, neurodegenerative diseases, and preclinical or impending disease.

First, individuals who have experienced traumatic brain injuries have been shown to be more inconsistent than healthy adults on cognitive tasks (Bleiberg, Garmoe, Halpern, Reeves, & Nadler, 1997; Collins & Long, 1996), particularly for those with greater cognitive complexity (Hetherington, Stuss, & Finlayson, 1996). Moreover, Hetherington and colleagues found individuals who were at earlier points in their recovery process (i.e., 5 years since the incident) were more inconsistent than individuals who were farther along in their recovery (i.e., 10 years). Further, the link between head injury and intraindividual variability was particularly pronounced when the frontal brain regions were damaged (Stuss, Murphy, Binns, & Alexander, 2003). Burton, Hultsch, Strauss and Hunter (2002) also found that individuals with head injury showed greater inconsistency in physical functioning across occasions than healthy adults, further supporting the hypothesis that intraindividual variability is a stable characteristic of an individual.

The neurological disturbance does not need to be due to a discrete insult however to be behaviourally associated with greater fluctuations in cognitive functioning. Rather, an abundance of evidence has shown that individuals with diseases that cause progressive neurological deterioration are more variable in their cognitive performance. A seminal study by Hultsch and colleagues (2000) compared the intraindividual variability in responding to RT tasks in three different groups of older adults: those with clinical dementia, those with arthritis, and healthy adults. The adults with dementia demonstrated

twice as much inconsistency as did the healthy and arthritic adults, who did not differ. Therefore, intraindividual variability appeared to be the result of endogenous neurological conditions, rather than exogenous somatic conditions like arthritis. Similar findings were reported by Strauss et al. (2002), who found individuals with dementia were more inconsistent in their physical functioning as well. It has since been demonstrated that the amount of intraindividual variability may vary even among neurological impairments. Burton, Strauss, Hultsch, Moll and Hunter (2006) investigated whether increased inconsistency was apparent regardless of the type of neurological condition, indicating potential causation by general neurological disturbance, or was exclusively or more strongly related to certain types of neurological disturbance, signifying variability's closer ties to a specific type of neurological insult. They compared older individuals with Parkinson's disease, Alzheimer's disease, and those with no neurological condition on 4 RT tasks. Consistent with previous research, both disease groups were more inconsistent than the healthy adults on all cognitive tasks, but the Alzheimer's individuals also displayed more variability than the Parkinson's patients.

Even stronger distinctions among the types of neurological disturbance have been made. Murtha, Cismaru, Waechter, and Chertkow (2002) found individuals with frontal lobe dementia showed greater amounts of inconsistency than individuals with Alzheimer's dementia. Relatedly, Walker et al. (2000) showed that patients with Lewy body dementia exhibited greater fluctuations in their cognition and attention than those with vascular or Alzheimer's dementia. Lewy body and frontal lobe dementia cause greater deterioration of the frontal lobes than Alzheimer's dementia (Hultsch et al., 2008), suggesting that inconsistency is predominantly linked to insult in the frontal brain

regions. Such a hypothesis fits with suggestions from the brain injury literature (Stuss et al., 2003).

Finally, given the demonstrations of increased intraindividual variability among those with neurodegenerative diseases, it has recently been investigated whether increased intraindividual variability may act as an early predictor of impending disease such as dementia. Mild cognitive impairment (MCI; Petersen et al., 1999) is proposed as an intermediary stage between normal functioning and dementia, where individuals perform below their age- and education-matched peers, but maintain otherwise normal functioning. This is hypothesized to represent a very early stage of impending dementia, at least in some cases (e.g., Albert, 2008). Thus far, all studies investigating this association have found that older adults with some form of MCI were more inconsistent than healthy older adults (Christensen et al., 2005; Dixon et al., 2007; Strauss, Bielak, Bunce, Hunter, & Hultsch, 2007). Further, individuals with multiple domains of cognitive impairment showed greater fluctuations in responding to cognitively challenging RT tasks than those with an isolated (single) area of impairment (Dixon et al., 2007; Strauss et al., 2007). Therefore, not only is increased intraindividual variability related to possible preclinical neurological disease, but it can also differentiate among subtle differences in neurological functioning.

Clearly, the neurological conditions described above are relatively stable characteristics of individuals, and their consistent positive links with inconsistency support the notion of reliability in within-person fluctuations in performance. The congruent findings between neurological conditions and increased intraindividual variability have also been key pieces of evidence regarding the hypothesis that intraindividual variability is a behavioural marker of underlying central nervous system

integrity. Specifically, increased inconsistency in performance appears to be indicative of brain disturbance or dysfunction (e.g., Hultsch & MacDonald, 2004; Li & Lindenberger, 1999). Further, the above research supports the association between intraindividual variability and the severity of neurological disturbance, where greater amounts of variability were observed with more severe brain disturbances, particularly those that affected frontal brain regions (e.g., Murtha et al., 2002; Stuss et al., 2003). Therefore, the degree to which an individual is inconsistent in their cognitive responding relative to healthy individuals may be a reliable measure of their neurological integrity.

Although the exact neurological cause of increased intraindividual variability is unknown, there are several possible determinants (MacDonald, Nyberg, & Bäckman, 2006). MacDonald and colleagues noted structural causes such as lesions to frontal gray matter, problems with neuronal connectivity because of white matter demyelization, and possible neuromodulatory changes. Li and Lindenberger (1999) showed that changing the gain parameter of catecholamines (neurotransmitters which affect the responsivity of neurons to incoming signals, effectively changing the neuronal signal-to-noise ratio) in neurocomputational simulations changed the resulting amount of inconsistency in cognitive performance. Most importantly, decreases in the gain parameter of the neurotransmitter, or decreased responsivity to neuronal signals, resulted in increased variability.

Despite the apparent viability of the hypothesis that intraindividual variability is a sensitive marker of neurological integrity, the above research is based on correlational and indirect associations with neurological disorders. Recently however, this hypothesis has shown further confirmation from brain-based research, and added to the verification of inconsistency's individual stability.

If intraindividual variability truly is a marker of neurological integrity, research that directly assesses the structural characteristics and activation of the brain should show significant associations with behavioural inconsistency. Further, the relationship should be in the expected direction and consistent with the hypothesis that increased intraindividual variability is maladaptive; thus, greater behavioural fluctuation in performance, more structural damage and maladaptive activation patterns. Although there are few studies that have investigated this association thus far, the evidence is positive. First, using functional magnetic resonance imaging (fMRI) on young to middle aged adults, Bellgrove, Hester, and Garavan (2004) found a significant association between intraindividual variability on a go/no-go response task and activation of the frontal cortex. As expected, increased inconsistency on the task was associated with poorer performance, which translated into poorer inhibition of the "go" response. However, rather than decreased brain activation, poorer and more inconsistent performers elicited more frontal brain activity in order to inhibit the target behaviour in the "no-go" situation. Next, studies have found significant links with structural brain characteristics as well. Bunce and colleagues (2007) completed fMRI scans on healthy older adults aged 60-64 years. They found frontal lobe white matter hyperintensities (WMH; i.e., white matter lesions that affect the efficiency of neuronal conduction) were significantly associated with intraindividual variability on speeded cognitive tasks. More importantly, this relationship was unique to WMHs in the frontal lobes, clearly demonstrating that deterioration of neural pathways in the frontal cortex plays a key role in increased intraindividual variability. However, although the association was significant, the resulting effect size and strength of the relationship appeared to be minimal. This was possibly due to the relatively good health of the participant sample. The final study by

Anstey et al. (2007) supports that possibility. Anstey et al. found only minimal associations between the size of the corpus callosum (CC) as detected by MRI scans and inconsistency on simple and choice reaction time in a healthy older sample, but significantly stronger relationships for those with a mild cognitive disorder (i.e., MCI; age-associated memory impairment or cognitive decline; mild neurocognitive disorder). Despite the inherent inclination to then propose that the mild cognitive disorder sample had more problems in the brain (in this case, a smaller CC area), and thus a greater link with inconsistency, there was no difference in the average CC size between the two groups. However, this finding can still be in accordance with the idea that intraindividual variability is an indicator of neurological disturbance. Anstey and colleagues proposed that the biological limits of brain reserve capacity must have been reached for the mild cognitive disorder group, or that they were functioning at their maximum ability, resulting in their stronger brain-behaviour relationship.

Although the association between brain structures and activation with behavioural inconsistency is not perfect or particularly strong, the significant associations serve as direct evidence that inconsistency in behaviour is linked to brain-based characteristics. In fact, Bunce and colleagues (2007) noted that it is likely the combination of a number of underlying neurological factors, including declines in catecholamines and structural changes that contribute to increased inconsistency. For example, Kelly, Uddin, Biswal, Castellanos, and Milham (2008) investigated the possibility that when completing a demanding cognitive task, individuals need to suppress the "default mode" network of brain regions which show activity when the brain is at rest (e.g., sleep) in addition to activating those required to complete the task. They hypothesized that poor suppression of the default mode network resulted in poor task performance and increased variability.

Using event-related fMRI scans, they computed the strength of individuals' negative correlation between their default mode and their task-positive (i.e., on task) networks, serving as an index of the degree of regulation of activity in those networks (i.e., a weak negative correlation indicated a greater likelihood that the two networks were simultaneously active whereas a strong negative correlation indicated adequate suppression of the default mode network while the task-positive network was active). Consistent with their expectations, Kelly et al. found that the strength of individuals' negative correlations was related to individual differences in variability in performing an arrow RT task. Therefore, it may also be the case that inconsistency in cognitive performance is related to an individual's ability to regulate competing neural processes, in addition to their structural characteristics and neuromodulatory mechanisms.

Overall, a substantial amount of research demonstrates that intraindividual variability in cognitive performance is a stable, endogenous characteristic of an individual. Inconsistency is of significant magnitude and stability relative to other types of variability (e.g., Hultsch et al., 2000; Nesselroade & Salthouse, 2004; Rabbitt et al., 2001), significantly related to level of ability in a variety of cognitive domains (e.g., Burton et al., in press; Hultsch et al., 2002; Jensen, 1992), and correlated with individual characteristics such as activity participation (Bielak, Hughes et al., 2007), risk of attrition (MacDonald et al., 2003), and even impending death (MacDonald et al., in press). An individual's inconsistency in cognitive functioning is also significantly related to their amount of inconsistency in other domains of functioning, such as sensorimotor performance (Li et al., 2001), physical ability (Strauss et al., 2002) and control beliefs (Bielak, Hultsch et al., 2007; Eizenman et al., 1997). Increased inconsistency is associated with a variety of neurological conditions, including traumatic brain injury

(e.g., Bleiberg et al., 1997; Stuss et al., 2003), dementia (e.g., Hultsch et al., 2000), Parkinson's disease (Burton et al., 2006) and symptoms consistent with mild cognitive impairment (e.g., Christensen et al., 2005). Behavioural inconsistency is also significantly related to brain activation (Bellgrove et al., 2004), structural brain characteristics (Bunce et al., 2007; Anstey et al., 2007), the availability of neurotransmitters (Li & Lindenberger, 1999), and the ability to regulate competing neural processes (Kelly et al., 2008). Further, the above findings have supported the proposal that inconsistency is a sensitive indicator of central nervous system disturbance (Li & Lindenberger, 1999). Clearly, inconsistency in cognitive performance is a worthwhile area of interest in developmental research. Interestingly, inconsistency is related to the normative aging process as well.

Age Differences and Changes in Intraindividual Variability

Cross-sectional Differences Across the Lifespan

A number of studies have found that intraindividual variability is notably higher at the two ends of the lifespan: childhood and older adulthood. Using cross-sectional data, Williams, Hultsch, Strauss, Hunter, and Tannock (2005) found cognitive inconsistency between the ages of 6 and 81 followed a U-shaped curve. Inconsistency was highest for the youngest age group, sharply decreased in adolescence and young adulthood, and slowly increased again through middle and older adulthood. Similar findings have been shown across the age ranges of 5-76 years (Williams, Strauss, Hultsch, & Hunter, 2007) and 6-89 years (Li, Lindenberger et al., 2004). However, the underlying mechanism of increased variability appears to be different at the two points in the lifespan: For the younger half of the sample, age accounted for a significant proportion of the variance in inconsistency in both the fast and slow ends of the RT distribution, and partialing the

effects of inconsistency in the faster RT trials partially reduced the age-related effects of inconsistency in the slow RT trials. The reverse was also true. Williams and colleagues (2005) argued that the mutual attenuation of the age-related effects by the slow and fast ends of the distribution indicated a general underlying process for inconsistency in childhood, in addition to specific influences unique to the faster and slower ends of the distribution. On the other hand, age differences for older adults were minimal in the fast end of the RT distribution relative to those in the slow end, and partialing inconsistency in the fast end of the distribution did not attenuate the age effect in the slow end. These results were consistent with proposals that inconsistency in older adulthood might be due to specific variability-producing processes such as attentional lapses (as evidenced by variability among the slower RT trials), rather than a general process such as slowing (e.g., Hultsch et al., 2002).

Reconciling these findings with the perspective that inconsistency is a marker of neurological integrity (e.g., Li & Lindenberger, 1999), it appears that as the brain develops in childhood, or is becoming differentiated and specialized, it is relatively unstable, and the greater fluctuations in children's cognitive performance reflect that. During older adulthood however, the brain appears to develop the opposite way, and undergoes dedifferentiation, or increasing generalization of the brain structures that control specific cognitive processes. The dedifferentiation hypothesis coincides with poorer cognitive performance with increasing older age, and more compressed functional organization of intellectual abilities (Li, Lindenberger et al., 2004). However, Li, Lindenberger and colleagues (2004) found that intraindividual variability is predictive of fluid intelligence and chronological age only in late adulthood and older age, and not in childhood. Therefore, intraindividual variability in cognitive performance may be

particularly informative of the integrity of the aging brain and impending aging-related disease. Further, if increased inconsistency suggests increased neurological disturbance, but also increases with older age, this measure may be particularly insightful in explaining normative cognitive aging.

In fact, greater variability in cognitive responding among healthy older adults compared to younger adults is well substantiated (e.g., Anstey, Dear, Christensen, & Jorm, 2005; Bunce, MacDonald, & Hultsch, 2004; MacDonald, Hultsch, & Bunce, 2006; Nesselroade & Salthouse, 2004). Hultsch and colleagues (2002) compared younger adults aged 17 to 36 years with three groups of older adults: young-old (54-64 years), mid-old (65-74 years), and old-old (75-94 years). The participants completed 4 RT tasks of varying cognitive difficulty, ranging from simple RT and choice RT to lexical and semantic decision. On all 4 tasks, older adults showed more trial-to-trial inconsistency in performance than the younger adults, particularly compared to those in the oldest age group (i.e., 75 years and up). Further, differences were apparent even within the older adult age range: old-old adults demonstrated significantly more intraindividual variability in cognitive performance than young-old (i.e., 55-64 years), and mid-old (i.e., 65-74 years) adults.

Interestingly, age differences in intraindividual variability are most apparent on speeded tasks that challenge cognitive functioning, such as those that place large demands on executive functioning or working memory. For example, Dixon and colleagues (2007) found significant age differences on all 3 RT tasks (simple RT, choice RT, and 1-back choice RT), but the greatest age effects occurred on the 1-back choice RT which required participants to ignore the present stimulus and instead respond to the target stimulus presented on the previous trial. Similar findings have been found for neurological

conditions, where the greatest distinctions among groups were on cognitively complex tasks (e.g., dementia versus healthy, Hultsch et al., 2000; dementia subtypes, Murtha et al., 2002; CIND status, Strauss et al., 2007). However, it does not appear to be the case that the complex cognitive tasks simply exacerbate the effects of the poorer motor and perceptual functioning which accompany normal older age. Bunce et al. (2004) demonstrated that greater variability in older adulthood was not due to slower motor processing, but rather caused by increased attentional and cognitive demands. Further, MacDonald, Hultsch et al. (2006) investigated whether decreased perceptual functioning in older age was the cause of greater inconsistency on complex RT tasks by asking if younger adults could be experimentally manipulated to show inconsistency levels on par with adults who were decades older. Visual degradation of the stimuli however, did not successfully age younger adults' performance to mimic that of older adults', supporting the proposal that greater inconsistency with age is due to a more general endogenous process.

Longitudinal Changes with Age

However, any true test of a developmental phenomenon requires longitudinal evidence. Because the above studies were cross-sectional, they only demonstrate that there are age *differences* in intraindividual variability in older adulthood, and not that intraindividual variability *increases* with older age, as such change statements require longitudinal data. To date, relatively few studies focusing on inconsistency have employed longitudinal techniques, and only two have investigated its relationship to other meaningful outcomes. However, the results are promising.

First, several studies have demonstrated that intraindividual variability does in fact increase with older age. An 8-year study by Fozard, Vercruyssen, Reynolds, Hancock,

and Quilter (1994) found within-person variability in responding to audio RT tasks increased over time, particularly for the more challenging task. A more thorough population-based study by Deary and Der (2005) compared the initial variability in RT performance of younger to middle-aged adults (16-63 years) to their performance 8 years later. Interestingly, they found variability in performance on simple RT remained stable up until around 50 years. In contrast, intraindividual variability on the choice RT task increased steadily from the mid-30s to the mid-60s. Therefore, in accordance with cross-sectional evidence (e.g., Strauss et al., 2007), complex RT tasks appear to be most attuned to subtle neurological changes that occur with age. Der and Deary (2006) replicated these findings with a different population-based study that spanned a much larger age range (18-94 years). Again, variability in performing the simple RT task did not increase until age 50, whereas fluctuations in performance on the complex RT tasks increased throughout the adult age range. Both studies also found women were more inconsistent than men on the choice RT task in mid-adulthood (i.e., 36-63; Deary & Der, 2005), and across the lifespan (Der & Deary, 2006), but this finding has been questioned. Reimers and Maylor (2006) examined RTs on a trial-to-trial basis (compared to the overall SD used by Der and Deary), and found the gender effect disappeared when the first two trials of the task were removed. That is, women were initially slower than men, but became faster than men across the remaining trials.

Next, two recent studies have not only corroborated that individuals become more inconsistent as they age, but also demonstrated the utility of intraindividual variability in predicting cognitive change. MacDonald and colleagues (2003) examined 6-year longitudinal change in inconsistency, and expanded their investigation to include whether intraindividual variability predicted corresponding change in level of cognitive

performance. Participants from the Victoria Longitudinal Study aged from 55 to 89 years (N=446) completed 4 RT tasks in addition to cognitive tests targeting processing speed, working memory, fluid reasoning, episodic memory, and crystallized verbal ability. Each test was completed at each of three waves over a 6-year period. First, there were marked increases in intraindividual variability, but only for the old-old participants (i.e., 75-94 years). The young-old (55-64 years) and mid-old (65-74 years) adults remained constant in their within-person variability or declined slightly over the 6 years. Therefore, in contrast to earlier findings (e.g., Der & Deary, 2006), inconsistency did not increase uniformly across the older age range. However, differences within the older adult range are consistent with cross-sectional findings that inconsistency is greater in the latter half of older adulthood (e.g., Hultsch et al., 2002). Further, 6-year change in inconsistency was greater on the verbal (i.e., lexical and semantic decision) compared to non-verbal RT tasks (i.e., simple and choice RT), which were arguably more cognitively complex and challenging. Next, MacDonald and colleagues investigated the longitudinal links between inconsistency and cognitive change. Impressively, the amount of inconsistency at the initial test occasion significantly attenuated 6-year decline on the cognitive tests by an average of 93%. In other words, cognitive inconsistency at time 1 significantly predicted cognitive change over the 6 years. However, do the *changes* in inconsistency also covary with the changes in cognition? Using hierarchical linear modeling, MacDonald and colleagues found significant associations between 6-year change in intraindividual variability and 6-year change in cognitive test score. On occasions where individuals were more inconsistent, they also tended to score lower on cognitive tasks completed at that wave of testing, independent of the average linear trend across time. The relationship between increasing intraindividual variability over time and declining

cognitive performance was significant for five out of the six cognitive measures (all except crystallized verbal ability). Interestingly, the significant covariation relationship was invariant across age groups, indicating the amount of cognitive decline per unit increase in inconsistency did not vary as a function of age. Therefore, although increases in inconsistency tended to be larger in old-old adulthood, the relationship between inconsistency and cognition remained stable across the older age range. These findings fit corresponding larger declines in cognitive ability past 75 years of age.

Lövdén, Li, Shing, and Lindenberger (2007) conducted a similar investigation with the old and very old (70-102 years) from the population-based Berlin Aging Study. They wanted to replicate the findings of MacDonald et al. (2003) with a much older sample and with five occasions of testing across a longer time frame (i.e., 13 years). They also wanted to extend these results by investigating the lead-lag relationship between inconsistency and cognition. Participants' level of cognitive ability was based on performance in two different cognitive domains: perceptual speed (assessed by digit letter) and ideational fluency (assessed by categories). Intraindividual variability was derived from another perceptual speed task (identical pictures) that recorded RT for each trial. Using latent growth curve modeling, Lövdén et al. found results consistent with those reported by MacDonald et al. (2003): individuals across the age range (70-102 years) became more inconsistent over time, with an average 1.2 unit increase every 2 years. The uniform increase in variability across the sample is consistent with the findings for the old-old group (75-94 years) in MacDonald et al. Lövdén and colleagues also found significant associations between longitudinal change in inconsistency and longitudinal changes in ideational fluency and perceptual speed, such that individuals who were more variable at one testing occasion tended to also score lower on the

cognitive tasks at that occasion. Further, the covariation relationship was significant across the entire range from 70-102 years, corroborating that the link between inconsistency and cognition remained stable in older adulthood (i.e., MacDonald et al., 2003). Overall, Lövdén and colleagues confirmed that significant covariation between cognition and inconsistency was apparent even among the old and very old in a population-based study, and more impressively, these results were not confounded by chronological age, suspected dementia, or time-to-death.

However, because random effects models are limited in their ability to address lead-lag relations as change and level are defined over the same period, and thus level does not precede change, it was unknown whether one variable caused change in the other. Did inconsistency signal subsequent negative longitudinal changes in cognitive performance, or did the relationship operate in the opposite direction, where cognitive decline occurred before participants showed more behavioural variability? To examine this further, Lövdén and colleagues employed bivariate dual change score models, which allowed for corresponding time-lagged associations between the variables and direct and simultaneous modeling of the variables' changes. The results showed greater inconsistency reliably preceded and predicted greater 2-year negative decline in ideational fluency (categories) and perceptual speed (digit letter), indicating that intraindividual variability could predict looming changes in level of cognitive performance. Further, the opposite model of level of ideational fluency did not influence later change in intraindividual variability. However, this same dissociation did not hold across cognitive domains; greater perceptual speed also predicted greater 2-year increased change in inconsistency. Regardless, in the first study to examine the lead-lag relationship between inconsistency in cognitive performance and cognitive level, there was evidence that

inconsistency may be "a developmentally early flag for impending old-age changes in mean levels of cognitive performance" (p. 2834), even if the strength of this predictive relationship might vary across cognitive domains. Further, this predictive possibility is particularly promising because the time-lagged relationship was apparent regardless of chronological age, suspected dementia, and time-to-death.

Overall, it is well documented that intraindividual variability is significantly associated with the aging process. Inconsistency in cognitive performance is positively associated with age, with older adults being more inconsistent than younger adults (Hultsch et al., 2008), and also significantly increasing with age, beginning around age 70 (e.g., Lövdén et al., 2007; MacDonald et al., 2003). More importantly, longitudinal studies of aging have verified the link between greater inconsistency and poorer cognitive ability over time by finding significant coupling relationships between change in intraindividual variability and change in cognition across 6 (MacDonald et al., 2003) and 13 years (Lövdén et al., 2007). Finally, Lövdén and colleagues (2007) identified the temporal order of this relationship among the old and very old adults as change in inconsistency preceding change in cognition, indicating that intraindividual variability in older adulthood may be a sensitive marker of impending cognitive decline or disease. This dissertation built on these findings by delving deeper into age-related change in intraindividual variability, its links within cognitive change, and whether intraindividual variability truly is indicative of later developmental outcomes.

Overview of Studies

A main goal of developmental research is to evaluate whether current outcomes can be predicted by certain preceding factors or variables. These types of studies are particularly valuable and informative of the developmental process, but few such studies

exist in the intraindividual variability literature. Longitudinal data is essential for this research area because intraindividual variability is believed to be an early behavioural marker of neurological disturbance and impending cognitive decline. Therefore, the only way to validate this hypothesis is through prospective longitudinal investigations. This dissertation presents two studies of intraindividual variability in a longitudinal context to further explore the relationship between short-term intraindividual variability and longer-term cognitive change in older adults. The first study addressed change in inconsistency over time, and the covariation of change in inconsistency with change in cognition. The second study addressed the capacity of intraindividual variability to predict later cognitive ability and other meaningful change outcomes such as cognitive status and attrition.

Study 1

There is substantial evidence that trial-to-trial fluctuations in cognitive performance increase as a result of the aging process. Specifically, intraindividual variability appears to linearly increase from early adulthood well into late older adulthood (e.g., mid-30s to mid-60s, Deary & Der, 2005; 18-94 years, Der & Deary, 2006), with an increased acceleration in old-old adulthood (e.g., 70-102 years, Lövdén et al., 2007; 75-89 years, MacDonald et al., 2003). Interestingly however, greater increases in intraindividual variability in old-old adulthood do not appear to change the underlying coupling relationship with cognitive performance. Rather, the relationship between cognition and variability appears to remain stable across old adulthood (i.e., 55-89 years, MacDonald et al., 2003), where the amount of cognitive decline per unit increase in inconsistency does not vary by age. Further, Lövdén et al. (2007) presented the first temporal evidence that inconsistency in cognitive performance significantly precedes cognitive decline, and may thus be able to predict impending cognitive changes in older

adulthood. However, because only two studies thus far have investigated the covariation relationship between inconsistency and cognition in older adulthood (i.e., Lövdén et al., 2007; MacDonald et al., 2003), it is unknown if the relationship will generalize to other time scales or age groups. Therefore, the longitudinal link between inconsistency and cognitive performance demands further replication and extension (Lövdén et al., 2007; Hultsch & MacDonald, 2004).

The age-related increase in inconsistency also appears to vary according to the cognitive difficulty of the task. For example, Deary and Der (2005) and Der and Deary (2006) found variability in responding to a simple RT task remained stable until age 50, whereas inconsistency in completing a choice RT task increased throughout the adult age range. Further, MacDonald et al. (2003) found 6-year change in inconsistency was greater on verbal (i.e., lexical and semantic decision) compared to non-verbal RT tasks (i.e., simple and choice RT). Similar findings have been found throughout the inconsistency literature, including inconsistency based on complex tasks providing greater sensitivity to cross-sectional age differences (e.g., Bunce et al., 2004; West, Murphy, Armilio, Craik, & Stuss, 2002), and other stable characteristics such as level and change of activity participation (Bielak, Hughes et al., 2007). Further, cognitively demanding tasks provided the best discrimination among varying degrees of mild cognitive impairment (Dixon et al., 2007; Strauss et al., 2007; but see Christensen et al., 2005). Therefore, it appears that the variability in responding while under high cognitive demand, compared to that observed while performing simple cognitive tasks, may be most attuned to the integrity of the neurological system, and thus subtle changes that occur with age. It remains to be seen however whether the difficulty of the inconsistency-based task affects the coupling relationship with cognitive performance as well.

The present study enhanced the inconsistency literature by addressing these issues. First, in terms of the generalizabilty of the longitudinal relationships, the sample was recruited on the basis of having concerns about their memory (rather than a normative aging study as used by MacDonald et al.), and covered the mid-old to old-old age range (i.e., 64-92 years). A shorter testing interval (i.e., yearly measurements) over 3 years was also used. Although MacDonald and colleagues used 6-year longitudinal data, because their participants were tested every three years, their covariation results were based on change relationships that occurred at 3-year intervals. Similarly, Lövdén et al. assessed the change relationship over 13 years, but because the participants were retested approximately every 2 years, the coupling of cognition and inconsistency was investigated at 2-year intervals. The present study was conducted over 3 years, and included 4 annual testing occasions. This shorter longitudinal interval will further explore the sensitivity of intraindividual variability to short-term cognitive changes, as it may be the case that the link between cognition and variability is stable only across time periods of certain length. Can inconsistency identify even annual developmental change? Finally, the present study used 10 RT tasks of varying complexity to assess whether the longitudinal relationship between intraindividual variability and cognition is stronger for tasks requiring greater cognitive effort. Lövdén et al. used only one type of trial-to-trial variability based on completing an identical pictures task (which is a form of choice RT) and there were significant associations with both assessed cognitive domains (perceptual speed and ideational fluency). On the other hand, MacDonald et al. analyzed 4 different measures of inconsistency (i.e., 2 non-verbal and 2 verbal RT tasks) with 6 different cognitive domains (e.g., working memory, fluid reasoning, episodic memory). The covariation relationship between inconsistency and cognition was significant in each

cognitive domain except crystallized verbal ability (vocabulary), but the authors did not indicate whether the relationship also varied by the complexity of the RT task (i.e., inconsistency). The present study provided a thorough examination of this question by analyzing inconsistency on 10 different RT tasks and assessing their relationship with 5 cognitive domains.

Research Questions

1. Does Wave 1 inconsistency significantly predict Wave 4 cognitive ability?

This initial question acted as verification of the longitudinal link between inconsistency and later cognitive ability in this sample. MacDonald et al. (2003) found that controlling for year 1 inconsistency significantly attenuated 6-year change in cognitive performance, suggesting similar associations may exist over a shorter 3-year time period.

2. Is there significant change in cognitive ability over a short time frame (i.e., 3 years?) Is there significant change in intraindividual variability over 3 years?

Both of these research questions looked to confirm that significant change did occur over the 3 years, and were investigated at both the group and individual levels. Previous investigations with this dataset have shown modest average gains in cognitive ability rather than decline. However, these analyses were based on shorter time frames (i.e., 1 and 2 years). Therefore, the overall pattern across the 4 measurement occasions was nonetheless expected to be one of decline, even if there was evidence of practice effects or improvement at the 2nd or 3rd time points. At the individual level, we expected significant between-person differences in cognitive change over time, with some individuals showing slight improvements, some maintaining their performance over time, and others showing decline in ability. For inconsistency, we expected to see an average

increase over the 3 years. Although inconsistency is hypothesized to be a stable trait of an individual, significant group-level variation in the amount of within-person variability has been found across 6 (MacDonald et al., 2003) and 8 years (Deary & Der, 2005). Despite the present study's relatively shorter time frame, the availability of four measurement points increased the likelihood of finding significant change in intraindividual variability over 3 years, and the potential for even small but significant variations in inconsistency change was maximized by the use of RT latencies. Based on the past significant group-level increases in inconsistency, we expected significant individual patterns in change in inconsistency as well.

3. Does 3-year change in cognitive ability covary with 3-year change in cognitive inconsistency?

Given the findings by Lövdén et al. (2007) and MacDonald et al. (2003) across 2 and 3-year periods respectively, it was reasonable to expect that these same patterns will be evident across a shorter time period. Specifically, we expected change in cognition to correspondingly change with inconsistency across the 3 years, such that responding at each wave of testing would be negatively correlated with the score on cognitive tasks (i.e., we would find individuals who were more inconsistent at one wave of testing would also score lower on cognitive tests completed at that same wave of testing). Therefore, we expected intraindividual variability to reflect even annual changes in cognitive ability.

4. Are there different relationships between intraindividual variability change and cognitive change based on the cognitive effort of the tasks?

Prior research suggests greater change in inconsistency may occur on complex RT tasks (e.g., Deary & Der, 2005), and this may in turn impact its relationship with

cognitive ability. We expected that tasks that involve more cognitive complexity would show stronger coupling relationships than those that require minimal cognitive effort.

Method

Participants

The final sample included 331 participants; however, due to various reasons including time and interest conflicts, 27 participants did not complete the first year of testing. Thus, data from a total of 304 community-dwelling older adults initially aged 64-92 years (208 women, 96 men) were analyzed in the study. All participants resided in the region of Victoria, British Columbia, Canada, and were participants of Project MIND (Mental Inconsistency in Normals and Dementia), a longitudinal research study in which participants are tested every year on an extensive battery of cognitive, physical, health, and psychological measures. The participants were originally recruited through advertisements in the local media (i.e., newspaper and radio) requesting healthy volunteers who were concerned about their mental functioning. Initial exclusionary criteria included a diagnosis of dementia by a physician or a Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) score less than 24, a history of significant head injury (defined as loss of consciousness for more than 5 minutes), other neurological or major medical illness (e.g., Parkinson's disease, heart disease, cancer), severe sensory impairment even with corrective aids (e.g., difficulty reading newspaper-size print while wearing glasses or hearing a normal conversation while wearing hearing aids), drug or alcohol abuse, a current psychiatric diagnosis, psychotropic drug use, and lack of fluency in English.

The participants were classified into two groups based on age at the first occasion of testing: a young-old group aged 64-74 years (n = 176, M = 70.09, SD = 2.91), and an

old-old group aged 75-92 years (n = 128, M = 80.06, SD = 4.01), of whom 18 participants were 85 years of age or older. Age group was included as a predictive factor in all proposed analyses as old-old adults have been shown to be more inconsistent that younger, young-old, and mid-old adults (Hultsch et al., 2002), and show larger increases in inconsistency over time (MacDonald et al., 2003).

Participants provided demographic and self-reported health information during an initial intake interview. In addition to the MMSE, several benchmark cognitive measures were obtained, including the Block Design and Vocabulary subtests of the Wechsler Adult Intelligence Scale-III (WAIS-III; Psychological Corporation, 1997), and the North American Adult Reading Test (NAART; Blair & Spreen, 1989). We computed estimates of full scale IQ (FSIQ) based on the age-adjusted Block Design and Vocabulary subtests (Sattler & Ryan, 1999) and premorbid IQ based on the NAART (Blair & Spreen, 1989).[1]

A 2 (age group) X 2 (gender) MANOVA (Multivariate Analysis of Variance) including all dependent variables was completed. There were significant multivariate effects for age group, $F(8, 293) = 5.49$, $p < .001$, $\eta^2 = .13$, and gender, $F(8, 293) = 2.50$, $p < .05$, $\eta^2 = .06$, but the Age Group X Gender interaction was not significant. There were significant age group differences for education, $F(1, 300) = 14.03$, $p < .001$, $\eta^2 = .05$; total chronic health conditions, $F(1, 300) = 8.89$, $p < .005$, $\eta^2 = .03$; MMSE, $F(1, 300) = 26.28$, $p < .001$, $\eta^2 = .08$; WAIS estimated FSIQ, $F(1, 300) = 5.58$, $p < .05$, $\eta^2 = .02$; and NAART estimated IQ, $F(1, 300) = 5.14$, $p < .05$, $\eta^2 = .02$. The young-old adults had significantly more education (M = 15.70, SD = 2.95) and reported fewer chronic

[1] Blair and Spreen's (1989) formula for estimating premorbid IQ using the NAART is $NAART_{estIQ} = 127.8 - 0.78\ (NAART_{errors})$. This formula is based on the Wechsler Adult Intelligence Scale – Revised, whereas our estimate of current IQ is based on the WAIS-III, which is somewhat more difficult. Thus, the discrepancy between the NAART estimate and the WAIS-III estimate will, if anything, slightly overestimate cognitive decline.

conditions (M = 2.56, SD = 1.77) than the old-old adults (education: M = 14.41, SD = 3.26; conditions: M = 3.42, SD = 1.99). In addition, the young-old adults performed significantly better on the MMSE (M = 29.03, SD = 1.01), and had higher estimated FSIQs (M = 120.66, SD = 11.86) and NAART estimated IQs (M = 117.04, SD = 6.62) than the old-old adults (MMSE: M = 28.35, SD = 1.30; FSIQ: M = 118.22, SD = 13.44; NAART estimated IQ: M = 115.69, SD = 6.87). There were no significant age differences for self-rated health measures, total chronic conditions, or total number of medications.

There were few significant gender effects. Women rated themselves in better health compared to a perfect state of health (M = 4.26, SD = .68) than men (M = 4.11, SD = .67; $F(1, 300) = 4.41$, $p < .05$, $\eta^2 = .01$), and also performed better on the MMSE (M = 28.82, SD = 1.13) than men (M = 28.57, SD = 1.4; $F(1, 300) = 5.93$, $p < .05$, $\eta^2 = .02$).

Overall, the participants were well educated (M = 15.16, SD = 3.14), ranging from 7 to 24 years of education, with only 10.2% having less than 12 years of schooling (n = 31). The participants were relatively healthy, with 65.5% having 3 or fewer medical conditions. The participants' perceptions of their relative health were also high, with 87.2% rating themselves good or very good compared to a perfect state of health, and 94.1% rating themselves good or very good compared to the health of other's their own age.

Procedure

Potential participants were initially screened for inclusion and exclusion criteria by a telephone interview. The various measures were administered during seven sessions (one group and six individual) scheduled over approximately 3 months. The entire testing

battery was repeated yearly 3 times, totalling 4 waves of data. The present study used the first 4 waves (years) of testing. The group testing session was held in our laboratories, and the individual testing sessions were conducted in the participant's home. Participants attended two testing sessions (one group and one individual) during which they provided demographic and health information, and completed tests used to assess cognitive ability. Participants then completed 5 individual testing sessions in the first wave of testing each scheduled approximately two weeks apart. In the second through fourth waves (years) of testing, only 4 individual testing sessions were completed. During each of these sessions, they performed the battery of reaction time tasks designed to assess short-term fluctuations in response speed. The tasks across the five (and four) testing sessions were identical, and the order of the tasks was invariant across participants. Because we were interested in variability, these five (and four) sessions were distributed across days of the week and times of the day rather than scheduling them at the same time.

Measures

Cognitive tasks

Cognitive ability was based on performance on 5 tests covering a range of cognitive domains. These tasks were administered only once per testing wave in a group format.

Perceptual speed. Perceptual processing speed was assessed using the WAIS-R Digit Symbol Substitution task (Wechsler, 1981). Participants were presented with a coding key pairing nine numbers (1 through 9) with nine symbols, below which were rows of randomly-ordered numbers with empty boxes. Participants were given 90 seconds to transcribe as many numbers to corresponding symbols as possible. The number of correctly completed items was recorded.

Reasoning. Reasoning ability was assessed using the Letter Series test (Thurstone, 1962). Participants were presented with sets of strings of letters which followed a particular pattern. Participants were required to provide the next letter in a target string of letters that was congruent with the pattern. Participants were given 6 minutes to complete twenty strings of letters. The outcome measure was the total number of items correct.

Episodic memory. The episodic memory construct was measured by a word recall task consisting of immediate free recall of 30 English words (Hultsch, Hertzog, & Dixon, 1990). The list contained 6 words from 5 taxonomic categories (e.g., birds, flowers) typed on a single page in unblocked order. Participants were given 2 minutes to study the list and 5 minutes to write their recall. The number of correctly recalled words was recorded.

Verbal fluency. Participants' verbal fluency was assessed using the Controlled Associations test (Ekstrom, French, Harman, & Dermen, 1976). The test required the generation of as many synonyms as possible to a set of four target words. Participants were given 6 minutes to complete the test. The outcome measure was the total number of correct synonyms.

Vocabulary. Vocabulary was measured by a 36-item multiple-choice test (Ekstrom et al., 1976), where the object of the test was to select the correct definition of a target word from five possible definitions. Participants were given 10 minutes to complete the test. The total number of correct items was recorded.

Reaction time tasks

The measurement of intraindividual variability was based on RT latencies from several multi-trial computer-based RT tasks. The tasks were of varying cognitive complexity. Participants responded to stimuli by pressing keys on an external keyboard

attached to a computer that was configured specifically for the task. For all tasks except for the finger tapping task, participants were instructed to emphasize speed in responding, and also attempt to minimize errors. These tasks were administered 5 times per testing wave (4 times in waves 2-4), approximately two weeks apart, in individually-administered sessions.

Finger tapping. Participants were instructed to tap a response key as rapidly as possible. Participants first completed the test with their left hand and then with their right hand. Each key tap resulted in a letter (L or R) appearing on the screen. One practice run of 24 key taps followed by a test run of 48 taps was administered for each hand. The latencies of each interkey interval for the left- and right-hand runs were recorded.

Choice reaction time (CRT). Participants were presented with a horizontal row of four plus signs, with a matching arrangement of keys on the response console. After a delay of 1000 ms, one plus sign changed into a box, and the participant was asked to press the key corresponding to its location as quickly as possible. The location of the box was randomly equalized across trials. 10 practice trials were first administered, followed by 52 test trials. The latencies for the 52 test trials were recorded.

Choice reaction time 1-back (BRT). This task used the same stimulus display and response keyboard as the basic CRT task. However, participants were instead instructed to press the key corresponding to the location of the box on the previous trial as quickly as possible. A total of 10 practice trials and 61 test trials were administered. Because participants made no response on Trial 1, the latencies of the remaining 60 test trials were used for analysis.

Arrow task. Left- and right-pointing arrows were presented along the medial-horizontal axis of the computer screen in one of three locations: left, center, and right. Only one

arrow was presented at a time. The response keyboard consisted of two keys corresponding to the left- and right-hand sides of the screen. Participants were instructed to press the key corresponding to the direction of the arrow as quickly as possible. There were three types of trials based on the combination of direction and location of the arrow stimuli: *congruent, incongruent*, and *neutral.* For *congruent* trials, the direction and location of the arrow denoted the same response (e.g., right-pointing arrow on the right side of the screen = press the right key). For *incongruent* trials, the direction and location of the arrow denoted the opposite response (e.g., right-pointing arrow on the left side of the screen = press the right key). For *neutral* trials, the arrow was presented in the center of the screen and had no apparent spatial displacement relative to the fixation point and the response keys. Each trial began with a fixation stimulus presented in the center of the screen for 250 ms. This was immediately followed by presentation of the arrow stimulus for 120 ms after which the screen cleared until the participant responded. Latencies were measured from the onset of the arrow stimulus until the response occurred. 10 practice trials and 100 test trials (40 congruent, 40 incongruent, and 20 neutral) were administered. Trial type occurred randomly across blocks of 10 trials so that trial type could not be predicted, but the proportion of trial types was consistent across the 100 trials. The trial by trial latencies for each of the three trial types were recorded.

Shape, Colour, & Task switching. Geometric figures varying in shape (square, circle) and colour (red, green) were presented in a white frame in the center of the screen. A task cue indicating the currently relevant stimulus dimension (shape or colour) was presented above the stimulus at the top of the frame. The response keyboard consisted of two keys. The right-hand key was to be pressed for circles and red objects and the left-hand key was to be pressed for squares and green objects. In the case of an error, the word error

appeared for 500 ms at the bottom of the frame. For each trial, the task cue word was presented 600 ms before the geometric figure stimulus. Participants were instructed to press the appropriate key as quickly as possible following presentation of the stimulus. Task cue and the stimulus object disappeared following the response, but the stimulus frame remained throughout. Three blocks of 50 trials were presented, each preceded by 10 practice trials. In the first block, participants were instructed to respond to the *shape* of the figure while ignoring the colour of the stimuli. In the second block, participants were instructed to respond to the *colour* of the figure while ignoring the shape of the stimuli. These single-task blocks were followed by a *task-switching* block in which the relevant response dimension (shape, colour) varied randomly without constraint. The trial by trial latencies for each of the three trial types were recorded.

Data Preparation

The distributions of raw latency scores were first examined for outliers at the level of individual trials. Extremely fast or slow responses most likely represent various sources of measurement error (e.g., accidental key press, distraction of participant). Lower bounds for legitimate responses were suggested by prior research (150 ms; Hultsch et al., 2002; Salthouse, Toth, Hancock, & Woodard, 1997), and were employed for all tasks except finger tapping, which had no lower limit for response time. Upper boundaries involved computing the mean and standard deviation for each task and occasion of measurement for each age group and removing any trials that exceeded the mean by three or more standard deviations. The average percentage of responses that exceeded these upper and lower boundaries across the 4 waves was: Left Tap, 1.8%; Right Tap, 2.0%; CRT, 1.2%; BRT, 3.1%; Arrow-congruent, 1.2%; Arrow-incongruent, 1.2 %; Arrow-neutral, 1.0%; Shape, 1.3%; Colour, 1.7%; and Switch, 1.9%.

Missing value estimates were imputed using a regression substitution procedure that forms individual equations of response times across all trials, which is then used to predict the missing RT entry (Hultsch et al., 2000). The procedures for eliminating outlying trials and imputing estimates for the missing values decreases within-subject variation, thus representing a conservative approach to examining intraindividual variability in response time performance.

Computation of intraindividual variability. Our measure of inconsistency was computed as the across-trial within-person individual standard deviation (ISD) about each individual's mean RT. Other techniques to calculate inconsistency have been statistically criticized, and our approach follows an alternative logic (see Hultsch et al., 2008). Potential confounding influences (e.g., age differences in mean RT; practice effects) and their higher-order interactions were partialed out using a split-plot regression:

Y = a + (age group)b + (trial)c + (age group X trial)d + e

By using the residuals (purified scores) produced from this regression, we effectively removed any systematic within- (i.e., trial) and between-subject (i.e. age group) sources of variance in mean RT, leaving only the possibility of evaluating each individual's unsystematic portion (i.e., inconsistency). These scores were then converted to standardized *T*-scores ($M = 50$, $SD = 10$) to enable comparisons across tasks, and each individual's SD was calculated.

Given the minimal number of errors on the RT tasks, ISDs were computed from all trials (correct and incorrect). Burton et al. (2006) demonstrated that the same pattern of results was obtained whether ISDs were computed from all trials or computed from only correct trials. ISD values were computed for each task, for each session, at each wave.

ISD Composites. In order to obtain the most reliable estimate of intraindividual variability at each wave, the ISDs were individually averaged across the 5 sessions for each task, producing one ISD score per task per wave for each individual (the ISDs were averaged across the 4 sessions for Waves 2-4). Although it could be argued that individuals may be most variable at the first session of testing due to their unfamiliarity with the task, the possibility exists that session 1 is not the most variable for each participant. Further, using only session 1 data represents a fairly liberal approach to testing the hypotheses (i.e., session 1 will provide us with the best chance of finding results). Rather, we argue that the most reliable measure of inconsistency is one that is averaged across the weekly sessions. In addition, session effects were not included in the present study because week-to-week variability is likely the result of different influences that those seen moment-to-moment (Hultsch et al., 2008). Week-to-week variability is more likely to be caused by external factors like stress or fatigue, while moment-to-moment variability is more likely to be caused by the internal brain-based influences that are the focus of the present study.

Next, ISD composites across the RT tasks were calculated based on factor structures reported in Strauss et al. (2007), using the same sample of participants. The 4 factors included: a Motor factor which consisted of the two tapping RT tasks (right and left), a Basic factor which involved the colour, shape, and CRT tasks, a Complex factor which consisted of the 1-back and switch RT tasks, and an Interference factor which included the three conditions of the arrow RT task (congruent, incongruent, and neutral). The total variance in the ISDs accounted for by the four factors was 82.15% (see Strauss et al., 2007). It was reasonable to expect a similar factor structure would exist over time, as the divergence of the factors made theoretical sense: the two tapping tasks were

expected to represent a similar motor construct; colour and shape were derivatives of the basic CRT task, and thus expected to represent a related construct to that assessed by CRT; the 1-back and switch tasks were the most demanding tasks as they required the participant to inhibit prior responses, and consequently expected to reflect comparable concepts; and the 3 conditions of the arrow task were expected to represent related information.

Statistical Analyses.

Multivariate multiple regression was used to ascertain a longitudinal link between Wave 1 inconsistency and Wave 4 cognitive performance at the group or average level. This analytic framework was used rather than advanced statistical techniques (e.g., multilevel modeling), because advanced techniques would have also focused on the group level (i.e., fixed intercepts) for the present question.

In contrast, Hierarchical Linear Modeling (HLM; also known as multilevel modeling, random coefficients modeling, or mixed modeling) was used to evaluate the rest of the longitudinal research questions, because the nature of the questions allowed them to substantially benefit from the statistical advantages of HLM in measuring change. First by focusing on change at the individual level, HLM allows two separate change questions to be asked: how does each person change over time (*within-individual change*), and what variables differentiate those individual patterns of change (*interindividual differences in change*). HLM does this by including two levels of statistical equations. At Level 1 or the within-subjects level, the equation is equivalent to a single regression equation for each individual case. The Level 2 equation, or between-subjects, evaluates whether stable group differences (e.g., age group, disease type) are associated with different trajectories of individual change across occasions. Next, cases

with missing data are not excluded with HLM, enhancing the statistical power to detect change. This is one of the key advantages for the longitudinal questions. HLM allows for individuals with even only one occasion of testing to be included in the analyses, by "borrowing strength" from other participants who have completed all waves of testing. For example, HLM will weight a participant who completed only one wave of testing less in the overall model compared to an individual who has data for all 4 time points. Further, HLM uses between-person comparisons to get a reasonable estimate of change for the person with one time point of data, allowing them to still be included in the analysis. It refers to participants similar to the individual with missing data, checks if those participants have more time points, and regresses the change of the participant with minimal data closer to the similar person with more reliable data. Finally, HLM provides more precise measurements of individual change because it does not require equal spacing between waves of testing. For example, although participants may have completed 4 testing waves, the amount of time between waves most likely varied from person to person. HLM can handle a variety of time intervals.

In relation to this final advantage of HLM, we calculated the following multiple indicators of time, which were later evaluated as to which provided the best fit to the data: Wave, which was constant across all participants (i.e., 0, 1, 2, 3); Time in study, which was 0 for each person at Wave 1, and subsequent entries were specific to the time in years since each individual's first wave of testing (e.g., 0, 1.23, 2.39, 3.44); and Chronological age, which was the person's exact age at each testing wave (e.g., 64.84, 66.01, 67.03, 68.59). Calculation of the last two time metrics were complicated by the fact that multiple testing sessions (i.e., group, intact, individual) were completed during each testing wave. Further, the 5 cognitive measures were completed at the first testing session

(i.e., group session), while the RT tasks were completed in the individual testing sessions. As a result, the date halfway between the first and last testing session for each wave [i.e., group session and 5th individual testing session (or 4th individual testing session in waves 2-4)] was decided upon as the time of testing for that wave[2]. Age at each testing occasion was subsequently calculated as the difference in days from the participant's birth date to their average testing date at each wave. Age was then centered at the age of the youngest participant at Wave 1 (64.48 years) in order to facilitate interpretation of the HLM parameters. The resulting exact ages at each testing occasion were then used to calculate the other individual-based time metric, time in study.

Results

The results are presented in four main parts. Each section builds on the prior analyses, and extends the investigation into the longitudinal relationship between cognitive ability and intraindividual variability in performance. First, the link between Wave 1 ISD composites and Wave 4 cognitive performance is established using group-based multivariate multiple regression. Next, individual-model based HLM more closely evaluates the longitudinal relationship by establishing whether significant change in cognition and inconsistency occurred over the short time period, and whether the changes covaried with one another across the 3 years. Finally, the possibility that cognitive complexity influences the short-term covariation relationship is examined using the HLM slope estimates.

Relation of Wave 1 Inconsistency to Wave 4 Cognitive Ability

[2] A handful of participants did not complete the group session or the 5th (or 4th) individual testing session for a particular test year. In such cases the next testing session (e.g., intake session; 4th session) was used to calculate their middle testing date for that wave.

Multivariate multiple regression was used to evaluate whether the Wave 1 ISD composites could predict cognitive performance level at Wave 4. The set of ISD predictors significantly predicted the set of 5 cognitive tasks, Pillai's trace = 1.39, $F(25, 1195) = 18.43$, $p < .001$. Individually however, only complex ISD was uniquely significant in predicting the set of 5 cognitive tasks [$F(5, 235) = 8.79$, $p < .001$, partial $\eta^2 = .16$], while basic ISD [$F(5, 235) = 2.23$, $p = .05$, partial $\eta^2 = .05$] and Interference ISD [$F(5, 235) = 1.97$, $p = .08$, partial $\eta^2 = .04$] were marginally significant. Motor ISD was not significant over and above the other predictors. Therefore, it appeared that the common influence of the ISD predictors was much stronger that any ISD composite on its own.

Table 1 shows that the score on each cognitive test 3 years later could be significantly predicted by the set of the Wave 1 ISD composites. The group of predictors accounted for the largest proportions of variance for the tasks requiring fluid intelligence, including letter series and digit symbol, and to a lesser extent, word recall. Relatively smaller, but significant amounts of variance were predicted for the similarities and vocabulary tasks. Therefore, a short-term longitudinal relationship (i.e., 3 years) between inconsistency and cognitive ability does exist, regardless of the type of cognitive task. Further, the direction of this relationship was almost exclusively negative [i.e., only for basic ISD with vocabulary was the beta weight positive (.086), but this was not significant], indicating that as ISD values increase, participants' cognitive performance decrease. However, the ability of each ISD composite to uniquely predict each cognitive task appeared to vary according to the complexity of the ISD composite.

Table 1

Univariate Test Results of Year 1 ISD Composites Predicting Year 3 Cognitive Test Score

	Cognitive Test				
	Digit Symbol	Letter Series	Word Recall	Similarities	Vocabulary
ISD Composite	*F*, partial η^2	*F*, partial η^2	*F*, partial η^2	*F*, partial η^2	*F*, partial η^2
Motor	ns	ns	ns	ns	5.04*, .02
Basic	9.53**, .08	ns	ns	ns	ns
Complex	17.77***, .07	34.56***, .13	11.43**, .05	7.60**, .03	ns
Interference	ns	8.68**, .04	ns	ns	ns
Overall	26.02***, .30	35.42***, .38	14.22***, .19	9.39***, .14	6.44***, .10

Notes. ISD composite: $F(1, 239)$; Overall: $F(4, 239)$, overall η^2. ISD = Intraindividual standard deviation.

*$p < .05$; **$p < .01$; ***$p < .001$.

Longitudinal Change in Cognition, Inconsistency, and their Covariation

Recent studies have demonstrated the variation in results that can occur depending on which time metric is employed in multilevel models. For example, Sliwinski, Hofer, Hall, Buschke, and Lipton (2003) found that time to dementia diagnosis was a better indicator of memory change than chronological age. Thus, before conducting the HLM models, it was necessary to first determine the appropriate time metric for the dataset. Table 2 provides a summary of the various time metrics with a sample of outcome variables. Because we are comparing time indices, and the models are not nested within one another, the Akaike's Information Criterion (AIC) was used to determine which model provided the best fit to the data. This metric penalizes overly complex models by multiplying the number of parameters by 2, and adding that value to the model deviance[3]. This allows non-nested models to be compared on a continuum scale, such that the lower the AIC value, the better the model fit (i.e., there is no critical value for a significant model difference).

For both word recall and vocabulary, the model using the time in study metric had the lowest AIC, and thus provided the best fit to the data. On the other hand, age provided the best model fit for Basic ISD. However, the requirement of age convergence was an inherent difficulty in using age as the time metric for our data. We did not expect the slope of cognitive ability or ISD to be the same across age due to well established findings that cognitive decline tends to occur at a faster rate in later older adulthood. If this was true in the present data, we could not use age as the time metric, which would assume that cognitive change was the same across age. We added age group as a Level 2

[3] Deviance is a numerical indication of how far the current model is from the best possible (saturated) model, for which the deviance value is 0.

Table 2

Model Fit Using Various Time Metrics for a Sample of Outcome Measures

Fixed effects	Slope (γ_{10})	t	AIC
Word recall			
Model 1: Wave	.29	3.92***	5534.76
Model 2: Time in Study	.29	4.17***	**5483.57**
Model 3: Age	-.09	-2.47*	5513.14
Vocabulary			
Model 1: Wave	.07	1.37*	4944.56
Model 2: Time in Study	.07	1.59*	**4897.06**
Model 3: Age	-.01	-.32	4900.25
Basic ISD			
Model 1: Wave	.04	1.31	3327.60
Model 2: Time in Study	.04	1.39	3323.94
Model 3: Age	.11	7.46***	**3286.26**

Notes. AIC = Akaike's Information Criterion (AIC). *$p < .05$; **$p < .01$; ***$p < .001$

predictor to the age model, and indeed found that the age groups differed in Basic ISD change over time, γ_{11}=.14, $t(302) = 3.75$, $p < .001$. Further, we failed to find age convergence for the other two outcomes as well: word recall, γ_{11}= -.21, $t(302) = -2.09$, $p < .05$; vocabulary, γ_{11}=-.17, $t(302) = 2.30$, $p < .05$. Because our final objective was to compare model results and provide an overall summary of change in cognition and intraindividual variability, it was necessary to choose one metric of time to be used for all outcome variables. Therefore, despite the variation in the time metric that provided the lowest AIC[4], we chose to use time in study as the time metric because of the failure to find age convergence for some variables.

3-Year Change in Cognitive Ability

To assess whether significant change occurred in each of the five cognitive measures, the following Level 1 and Level 2 equations were used

$$\text{Level 1: Cognitive Measure}_{ij} = \beta_{0j} + \beta_{1j}(\text{Time in Study}) + e_{ij}$$

$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

where the Level 1 equation examined individual rates of change across each individual's time in study. Specifically, the change in cognitive performance for a given individual (i) at a given wave (j) was a function of that individual's cognitive performance at the first wave of testing (β_{0j}; the intercept), plus that individual's average rate of change in cognitive performance across time in study (β_{1j}; the slope), plus an error term reflecting within-subject residual variance remaining to be explained after controlling for time in study (e_{ij}; the deviation from their own individual best-fitting regression line). At Level

[4] It was not the case that all cognitive measures favoured time in study, and all ISD measures favoured age. Other than wave providing the worst model fit across all outcome variables, there was no meaningful pattern across the outcomes.

2, or the between-subjects level, the intercept (β_{0j}) for each individual was modeled as a function of the mean sample cognitive performance for Wave 1 (γ_{00}), plus random error in estimating between-individual differences in intercept (u_{0j}). The slope (β_{1j}) for each individual was modeled as the average sample cognitive change per year increase of being in the study (γ_{10}), plus random error in estimating between-individual differences in slope (u_{1j}).

We found that for the majority of the cognitive tasks, there was an average small positive slope over the 3 years (γ_{10}; see Table 3). For example, with each additional year of being in the study, the average participant recalled .29 more words on the recall task. Only vocabulary performance did not significantly change over time. However, there was evidence across all tasks of significant individual differences in both cognitive ability at the first wave (u_{0j}), and change in cognitive ability over the 3 years (u_{1j}). For example, Figure 1 demonstrates the substantial between-person variability in word recall over time for a random sample of participants. There was also a substantial amount of variance in changes within each individual (e_{ij}; fluctuations from their own regression line) that remained to be explained after accounting for time in study.

Given the wide range of ages in the sample, we also investigated whether age group could significantly account for some of the between-person differences in cognitive change. Age group was added to the model as a Level 2 predictor, resulting in the following equations

$$\text{Level 2:} \qquad \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Age Group}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{Age Group}) + u_{1j}$$

As a result, each individual's intercept (β_{0j}) was now estimated as a function of the average starting point for the young-old age group (γ_{00}), plus the difference in average

Table 3

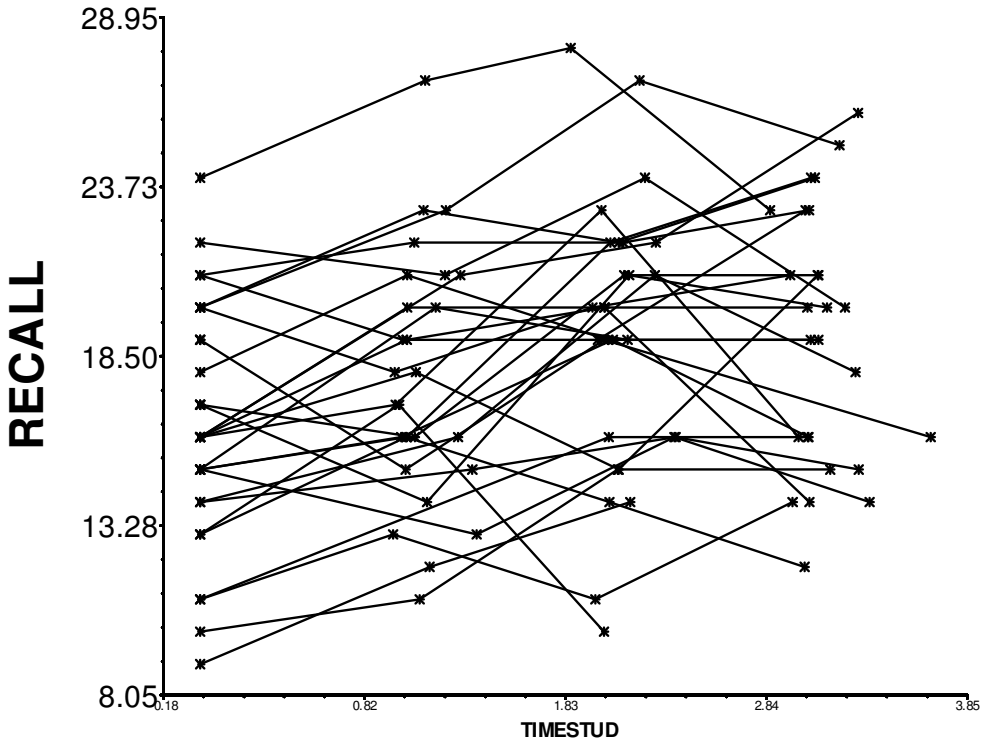
3-Year Change in Cognitive Performance

Cognitive Task	Fixed Intercept (γ_{00})	Fixed Slope (γ_{10})	Within-person Error (e_{ij})	Between-person Error in Intercept (u_{0j})	Between-person Error in Slope (u_{1j})
Digit Symbol	42.08***	.49**	22.62	89.45***	2.48***
Letter Series	9.10***	.13**	3.27	19.79***	.12*
Word Recall	16.52***	.29***	4.78	12.90***	.37***
Similarities	13.33***	.22*	9.16	19.80***	.41*
Vocabulary	30.03***	.07	2.17	16.49***	.15***

Notes. γ_{00} refers to the average cognitive performance at the first wave of testing. γ_{10} refers to the average amount of cognitive change per year increase of being in the study. e_{ij} refers to the random error in within-person estimation that remains after accounting for time in study; there are no significance tests for this effect. u_{0j} describes the random variance or between-person differences in the intercept or starting point of cognitive ability. u_{1j} refers to the between-person variance in slope of cognitive ability over time. *$p < .05$; **$p < .01$; ***$p < .001$.

Figure 1

Word Recall Change Over Time in Study for a Sample of Participants



intercept between the age groups (γ_{01}) which was multiplied by which age group the individual was in (young-old = 0; old-old =1), plus random variance in estimating between-individual intercepts (u_{0j}). Correspondingly, each individual's slope estimate (β_{1j}) was now a function of the average change for the young-old age group (γ_{10}), plus the difference in average slope between the two groups (γ_{11}) which was multiplied by which age group the individual was a member of, plus random error in estimating between-person differences in slopes (u_{1j}).

As expected, there were significant age group differences in the starting value (γ_{01}) on all cognitive tasks; as a group, the young-old adults achieved higher cognitive scores than the old-old adults at Wave 1. The parameter of interest however, was whether there were age group differences in cognitive change or slope (γ_{11}). Surprisingly, there were limited significant effects across the tasks. For word recall, the average young-old participant remembered .43 more words each year of being in the study ($p < .001$), while the average old-old participant recalled only .09 more words per year ($p_{diff} < .05$). The age group differences for vocabulary were more pronounced, as the average young-old adult increased their score .17 points per year ($p < .01$), but the prototypical old-old adult declined .08 points with each additional year of being in the study ($p_{diff} < .05$). Conversely, there were no significant group differences in slope for digit symbol, letter series, and similarities. Therefore, age group did not make a significant difference in the rate of cognitive change over the 3 years for those tasks.

Overall, the significant random effects within-individuals (i.e., in each individual's own regression line), and between-individuals in intercept and slope suggest that there is significant variance remaining to be explained by other Level 1 and Level 2 predictors (e.g., intraindividual variability).

3-Year Change in Intraindividual Variability

To assess whether significant change occurred in each of the four intraindividual variability measures (ISD composites), the following Level 1 and Level 2 equations were used

$$\text{Level 1: ISD Composite}_{ij} = \beta_{0j} + \beta_{1j}(\text{Time in Study}) + e_{ij}$$

$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

These models are identical to those used to investigate 3-year change in the cognitive tasks, substituting ISD composite as the outcome measure. Consequently, refer to the cognitive change section for a more thorough description of the equations.

Table 4 shows that slope or change in the amount of motor, basic, and complex inconsistency (γ_{10}) failed to significantly differ from zero over the 3 years. Therefore, as a group, the participants remained remarkably stable in their amount of inconsistency across the short time period. They did however show a small positive increase in their inconsistency on the interference tasks. For each year of being in the study, the average participant became .10 standard deviations more inconsistent in their performance on the interference tasks. All four ISD composite had significant random effects, indicating substantial between-person differences in the intercept (u_{0j}) and slope (u_{1j}) across time. For example, Figure 2 demonstrates the variation in change of complex ISD over time. Therefore, considerable variation existed around each ISD composite's best fitting regression line. This was also true for those ISD composites which did not show significant average change in ISD, indicating that not all participants remained stable in their intraindividual variability over time. Further, there was significant within-person error (e_{ij}; fluctuations from their own regression line) that remained to be explained after

Table 4

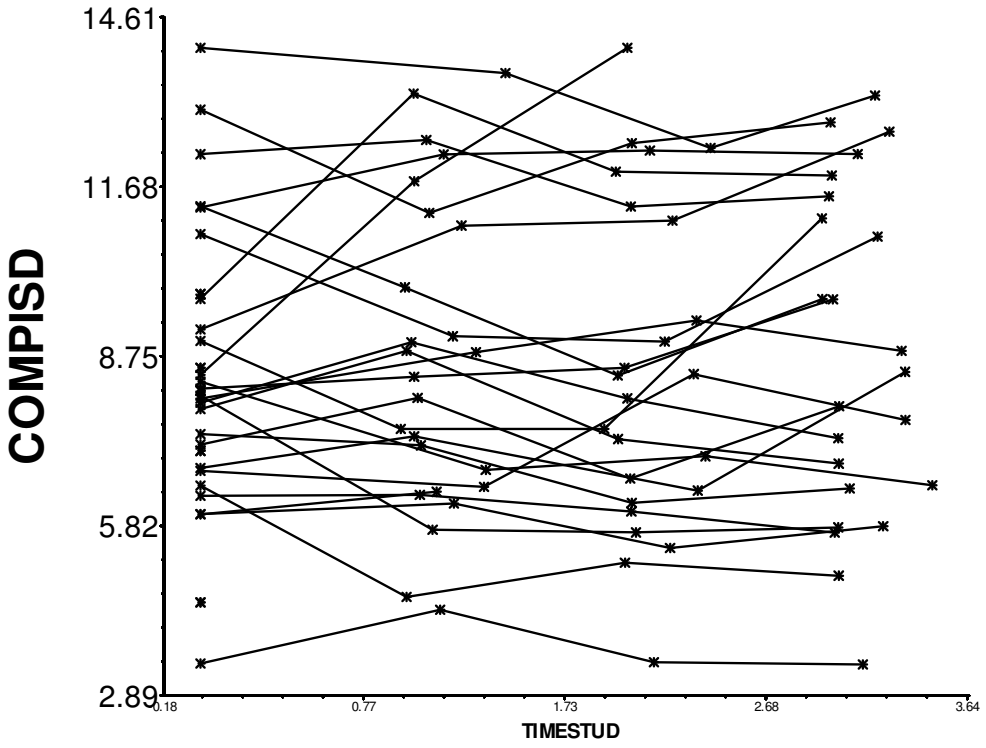
3-Year Change in Intraindividual Variability

ISD Composite	Fixed Intercept (γ_{00})	Fixed Slope (γ_{10})	Within-person Error (e_{ij})	Between-person Error in Intercept (u_{0j})	Between-person Error in Slope (u_{1j})
Motor	5.49***	.01	1.04	4.38***	0.16***
Basic	7.75***	.04	.45	3.09***	0.09***
Complex	7.62***	-.02	.48	6.15***	0.18***
Interference	7.25***	.10**	1.06	7.09***	.20***

Notes. γ_{00} refers to the average amount of ISD at the first wave of testing. γ_{10} refers to the average amount of ISD change per year increase of being in the study. e_{ij} refers to the random error in within-person estimation that remains after accounting for time in study; there are no significance tests for this effect. u_{0j} describes the random variance or between-person differences in the intercept or starting point of intraindividual variability. u_{1j} refers to the between-person variance in slope of intraindividual variability over time. *$p < .05$; **$p < .01$; ***$p < .001$.

Figure 2

Complex ISD Change Over Time in Study for a Sample of Participants



accounting for time in study.

Following from the cognitive change analyses, we also investigated whether age group could significantly account for some of the between-person differences in intraindividual variability. Age group was added to the model as a Level 2 predictor, resulting in the following equations

$$\text{Level 2:} \qquad \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Age Group}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{Age Group}) + u_{1j}$$

Again, these equations are identical to those analyzing the effect of age group on 3-year cognitive change, and a more detailed explanation of these equations can be found there.

Consistent with expectations, the old-old adults demonstrated higher starting rates of inconsistency across all 4 ISD factors than the young-old adults (γ_{01}). However, we were particularly interested in whether age group would affect the rate of change in inconsistency over time (γ_{11}). Although there were no significant differences among the groups for motor or interference ISD, the young-old participants were significantly more consistent than the old-old participants on basic ISD and complex ISD. For basic ISD, the average young-old participant experienced no change across the 3 years, but the average old-old participant was .17 standard deviations more variable with each passing year of being in the study ($p_{\text{diff}} < .001$). For complex ISD, the absolute difference between the age groups was similar to that for basic ISD, but the average young-old actually showed a .09 decrease in intraindividual variability ($p < .05$), while the average old-old participant became .07 units more variable with each year of being in the study ($p_{\text{diff}} < .05$). Overall, the age group effect on 3-year change in inconsistency was modest, and limited to higher complexity RT tasks.

Association Between 3-Year Change in Cognitive Performance and Inconsistency

To assess whether significant covariation existed between 3-year change in cognitive performance and inconsistency, the following Level 1 and Level 2 equations were used

Level 1: $\text{Cognitive Measure}_{ij} = \beta_{0j} + \beta_{1j}(\text{Time in Study}) + \beta_{2j}(\text{ISD Composite}) + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$

$\beta_{2j} = \gamma_{20} + u_{2j}$

where the Level 1 equation examined individual rates of change in cognitive performance in relation to individual time in study and rate of change in inconsistency. Because the previous analyses investigating 3-year cognitive change by time in study showed significant random effects, we chose to keep this parameter (u_{1j}).

However, the prior analyses inquiring into whether significant change occurred in inconsistency did not provide insight into whether the random effect of ISD (u_{2j}) would be significant in relation to cognitive change. In fact, preliminary analyses using the above covariation model across the 20 different combinations of the 5 cognitive tasks with each of the 4 ISD composites found only 6 instances of the random ISD effect being significant or marginally significant. To investigate this further, the random time in study effect (u_{1j}) was fixed in each model to determine if the ISD random effect (u_{2j}) would consequently emerge significant. However, this was not the case for the majority of the models (18/20), and instead often caused the significance of the random ISD effect to decline even greater to $p > .5$. Further, the variance associated with individual differences in time in study was always substantially greater than the individual differences associated with the ISD composite (e.g., σ^2 time in study = 0.15 versus σ^2 ISD = .02). Combined with the knowledge that the 20 various models needed to be similar to allow

model comparison, we subsequently chose to fix the ISD composite slope, resulting in the following Level 2 equations

$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

Therefore, we chose to remove the possibility of individual differences in the covariation relationship between cognitive change and inconsistency. Only the group or average relationship (i.e., fixed effect) would be estimated.

Specifically, the change in cognitive performance for a given individual (i) at a given wave (j) was a function of that individual's cognitive performance at the first wave of testing (β_{0j}; the intercept), plus that individual's average rate of change in cognitive performance with each additional year of being in the study independent of inconsistency performance (β_{1j}; Slope 1), plus that individual's average rate of cognitive change with each unit (i.e., standard deviation) increase in inconsistency independent of time in study (β_{2j}; Slope 2), plus an error term reflecting within-subject residual variance (i.e., deviations from their individual regression line) remaining to be explained after controlling for time in study and inconsistency (e_{ij}). There were three corresponding Level 2 equations. The intercept (β_{0j}) for each individual was modeled as a function of the mean sample cognitive performance for Wave 1 (γ_{00}) when inconsistency was zero, plus random error in estimating between-individual differences in intercept after accounting for both time in study and inconsistency (u_{0j}). Slope 1 (β_{1j}) for each individual was modeled as the average sample cognitive change per year increase of being in the study controlling for inconsistency (γ_{10}), plus random error in estimating between-individual differences in slope after accounting for time in study and

inconsistency (u_{1j}). The final Level 2 equation estimating each individual's predicted rate of cognitive change per unit increase in inconsistency (β_{2j}) was based on the mean sample's average amount of change in relation to the two measures while controlling for time in study (γ_{20}). This slope parameter was of primary interest, as it indicated whether for example, cognitive ability was lower when inconsistency in performance was higher.

Tables 5a) and b) show that the overwhelming majority of slope estimates between inconsistency and cognitive change were negative, in that for each one unit increase in inconsistency, participants' cognitive performance correspondingly declined, controlling for time in study. For example, for each additional unit increase in intraindividual variability on the Basic RT measures (i.e., 0.1 standard deviation), the average participant produced .58 less correct words on the similarities task. Further, the direction of the covariation relationship was identical across the continuum of cognitive tasks and inconsistency measures; more inconsistency, lower cognitive score. In addition, the negative covarying relationship between cognitive performance and ISD was invariant across time; regardless of the time in study, at testing occasions when a participant's variability increased, their cognitive performance also decreased. For example, Figure 3 illustrates how changes in Basic ISD were related to changes in digit symbol performance for all participants. These changes were evident at any time point in the study.

Potential sources of between-subject variance in terms of age grouping were investigated next. However, the intention to later compare the covariation relationship between cognitive performance and inconsistency across tasks and levels of cognitive complexity made it again necessary to decide on identical Level 2 equations; Specifically, whether age group should be included in the Level 2 equations for the intercept and slope for time in study. It is well established that there are cross-sectional cognitive differences

Table 5a)

Rate of Cognitive Change as a Function of Change in Inconsistency, Controlling for Time in Study (Digit Symbol, Letter Series, Word Recall)

	Slope (γ_{20})	SE	t
Digit Symbol			
Motor	-.51	.14	-3.67***
Basic	-2.32	.17	-13.58***
Complex	-1.77	.14	-12.16***
Interference	-1.01	.12	-8.13***
Letter Series			
Motor	-.20	.05	-3.80***
Basic	-.51	.07	-7.18***
Complex	-.63	.05	-11.54***
Interference	-.30	.05	-6.21***
Word Recall			
Motor	-.20	.06	-3.31**
Basic	-.57	.08	-7.30***
Complex	-.53	.06	-8.86***
Interference	-.35	.05	-6.59***

Notes. γ_{20} values reflect the average amount of change in the cognitive task per one unit increase in the inconsistency composite controlling for the average time in study effect. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 5 b)

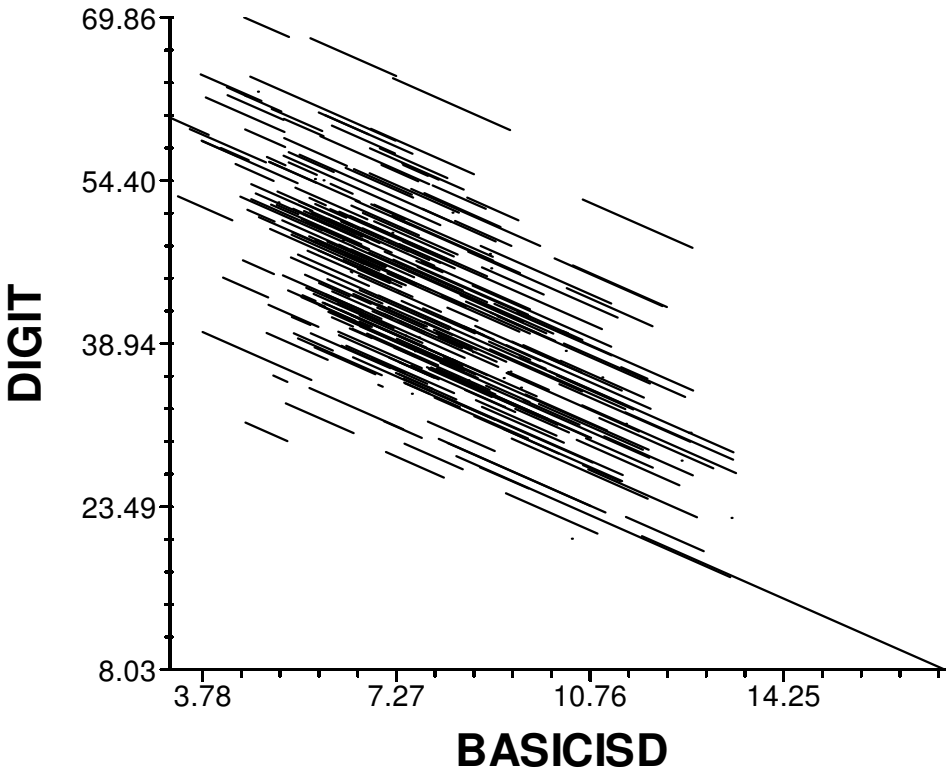
Rate of Cognitive Change as a Function of Change in Inconsistency, Controlling for Time in Study (Similarities, Vocabulary)

	Slope (γ_{20})	SE	t
Similarities			
Motor	-.06	.08	-.78
Basic	-.58	.10	-5.68***
Complex	-.58	.08	-7.42***
Interference	-.29	.07	-4.13***
Vocabulary			
Motor	-.06	.05	-1.31
Basic	-.14	.06	-2.18*
Complex	-.21	.05	-4.14***
Interference	-.04	.04	-1.00

Notes. γ_{20} values reflect the average amount of change in the cognitive task per one unit increase in the inconsistency composite controlling for the average time in study effect. *$p < .05$; **$p < .01$; ***$p < .001$.

Figure 3

Estimated Individual Regression Lines Describing Change in Digit Symbol and Change in Basic ISD at any Time in Study



even within the older adult age range, such that old-old adults tend to perform more poorly on cognitive tasks than young-old adults. Thus, it was reasonable to expect and subsequently allow potential age group effects in the intercept. In terms of age group effects for time in study, the previous analyses (see cognitive change across 3 years) indicated that there were no age group differences in change on the letter series, similarities, or digit symbol tasks. However, there were significant age group effects for the word recall and vocabulary tasks. Preliminary analyses comparing models with age group predicting all three Level 2 equations (i.e., intercept, time in study, and ISD) with those that only included age group for the intercept and ISD Level 2 equations showed little impact on the overall model. Some ISD slopes showed modest changes (e.g., γ_{20} = -.53 with age group at time in study vs. γ_{20} = -.54 without age group at time in study), but the significance of the effect did not change. Consequently, we chose to not include age group as a potential Level 2 predictor for the slope of time in study (β_{1j}). Rather, age group would only be a predictor of between-person differences for the intercept of time in study (β_{0j}), and for the slope of the ISD composite (β_{2j}).

The resulting equations evaluating between-subject differences in the covariation relationship between each cognitive task and each ISD composite were the following

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Age Group}) + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$

$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{Age Group})$

Of particular interest, each individual's slope estimate (β_{2j}) was now a function of the average ISD change for the young-old age group controlling for time in study (γ_{20}), plus the difference in ISD slopes between the two groups (γ_{21}) which was multiplied by which age group the individual was a member of (young-old = 0; old-old =1).

The models were remarkably consistent in indicating that there were no age group differences in the covariation between ISD and cognitive change (i.e., $p > .05$ for γ_{21}). Thus, the young-old adults experienced the same unit decrease in their cognitive score with each unit increase in inconsistency as the old-old adults. There were however two instances of significance: for the typical young-old adult, a one unit increase in motor ISD resulted in a lower letter series score by .25 ($p < .01$), whereas there was little change for old-old adults ($p_{diff} < .05$); and the average young-old adult's vocabulary score actually increased by .18 with each standard deviation increase of interference ISD ($p < .05$), whereas the average old-old adult's decreased by .14 ($p_{diff} < .01$).

Variations in the Covariation Relationship due to Cognitive Complexity

Tables 5a) and b) demonstrate that although the significant covariation relationships between cognitive change and inconsistency change are constant in their direction, controlling for time in study, the magnitude of the relationship varies according to the cognitive complexity of the paired measures. Further, these differences appear to be influenced by the difficulty of both the cognitive task and the RT measures used to calculate intraindividual variability. For example, the repercussions of a one unit increase in interference ISD were much greater for performance on the digit symbol task (i.e., 1.01 unit decline), than it was for completing the letter series task (i.e., .30 unit decline). Similarly, showing greater inconsistency on the motor tasks resulted in less corresponding cognitive change (.06 to .51 fewer points), compared showing the same one unit increase in inconsistency on the complex ISD measures (.21 to 1.77 fewer points).

To investigate the differences in ISD complexity further, we conducted pairwise comparisons of the fixed ISD slope estimates (γ_{20}) in the covariation models. Specifically, we expected the complex tasks to show stronger covarying relationships.

Thus, we evaluated whether the complex ISD slope for each cognitive task was significantly different from the corresponding motor, basic, and interference models.

The following formula was used

$$\frac{\text{Complex } \gamma_{20} - \text{Other } \gamma_{20}}{SE_{difference}} \rightarrow \frac{\text{Complex } \gamma_{20} - \text{Other } \gamma_{20}}{\sqrt{(\text{Complex } SE^2 + \text{Other } SE^2)}}$$

The strongest differentiations by ISD complexity were on the letter series and digit symbol tasks, where complex slope estimates were significantly stronger than the motor ($p < .001$) and interference ($p < .001$) ISDs. Significant differences between the complex and motor estimates were also found for similarities and word recall ($p < .001$), but the differences between the complex and interference ISDs were smaller for the similarities task ($p < .01$), and only significant at $p < .05$ for word recall. The poorest differentiation was on the vocabulary task, where only the complex and interference ISD estimates significantly differed ($p < .01$).

Despite these findings, the complex ISD estimates were not significantly different from the basic ISD estimates ($\alpha = .01$ for a one-tailed test) for any of the 5 cognitive tasks. Therefore, it appeared that both the basic and complex ISD composites provided the strongest covariation relationships with cognitive performance. To investigate this further, we also compared the slope estimates from basic ISD with those from interference and motor ISD. Although basic ISD performed to the same level as the complex ISD on the digit symbol and word recall tasks, differences were evident on the rest of the cognitive measures. Unlike complex ISD, basic ISD was not significantly different from the interference ISD estimates for letter series, similarities, and vocabulary. Therefore, both complex and basic ISD had superiority over the simplest ISD composite,

motor ISD, but complex ISD provided the strongest estimate over interference ISD across all cognitive measures.

For interest's sake, we also compared the motor and interference ISD estimates across the tasks. Surprisingly, interference ISD was superior to the motor ISD estimate on only the digit symbol task (p <.01), and the two ISD composites did not differ on the rest of the cognitive tasks.

Discussion

The purpose of the present study was to replicate and extend current knowledge of the longitudinal nature of inconsistency and its covariation relationship with cognition in older adulthood. Specifically, it investigated the short-term changes in inconsistency and cognition, whether inconsistency could significantly predict cognitive ability 3 years later, and the sensitivity of the 1-year coupling change relationship between the two measures. This investigation also addressed whether the cognitive complexity of the inconsistency-based and cognitive tasks affected the strength of the coupling relationship.

Predicting Level of Cognitive Ability 3 Years Later

Consistent with expectations, individuals' initial level of inconsistency in performing the RT tasks significantly predicted their level of cognitive ability 3 years later. These results corroborated previous research showing significant prediction 6 years later (MacDonald et al., 2003), but were the first to provide evidence of a short-term (i.e., 3-year) longitudinal link between inconsistency and cognition. In the multivariate multiple relationship, the combination of all four Wave 1 ISD composites significantly predicted the sample's variance in the five Wave 4 cognitive scores. The overlap of the different measures that were also based on different types of tasks (i.e., RT versus accuracy-based) suggests that both sets of measures may be dynamically linked, either by

one another, or by the same underlying process. Further, the fact that the overlap was found in a 3-year predictive relationship (i.e., inconsistency at Wave 1 and cognition at Wave 4) indicates that there is substantial long-term stability in these underlying mechanisms that are characteristic of an individual. Together with the finding that nearly all predictive relationships were in the expected negative direction (i.e., more inconsistent responding at Wave 1 was related to poorer cognitive ability 3 years later), these results are congruent with the hypothesis that inconsistency is a behavioural manifestation of neurological integrity, and may be an early marker of change in cognitive ability (e.g., Hultsch et al., 2000; Lövdén et al., 2007).

The group of ISD composites predicted the largest proportions of variance on the letter series, digit symbol, and word recall tasks, and relatively smaller, but still significant amounts on the similarities and vocabulary tasks. Therefore, inconsistency may be a stronger predictor of cognitive domains that rely on fluid compared to crystallized intelligence. The difficulty of the inconsistency-based RT tasks also impacted the predictive relationship, with the complex ISD composite showing the strongest unique prediction of later cognitive ability. These results suggest similar task-related differences may be evident in the coupling relationship between the two measures.

3-Year Change in Cognition and Inconsistency

A more reliable test of the hypothesis that inconsistency could be an indicator of cognitive change was whether the two measures actually changed together over time, or were coupled, in that change in one resulted in corresponding change in the other. Before we could investigate this possibility however, we had to ensure that significant change in cognition and intraindividual variability did in fact occur over the short time period.

Cognition

The average change (i.e., fixed slopes) for nearly all cognitive domains was unexpectedly positive, indicating that the average participant improved on the various cognitive tasks with each additional year of being in the study. Only vocabulary performance did not significantly change over time. Cognitive tasks that rely on fluid and processing abilities tend to decline with age (i.e., speed, reasoning, memory), while crystallized abilities (i.e., verbal ability) appear to be relatively unaffected until later older adulthood (Salthouse, 2004). Therefore, only the present results for vocabulary were consistent with previous research. In fact, MacDonald and colleagues (2003) used a sample recruited from the same population, and across a similar age range (i.e., 55-89 years), but found significant average decline over 6 years in comparable cognitive domains (i.e., processing speed, working memory, fluid reasoning, episodic memory, and crystallized verbal ability). Therefore, the most probable cause for the difference in results was the short time frame of the present study; 3 years did not appear to be enough time for significant average cognitive decline to occur. Additionally limiting the likelihood of observing significant group decline was that the sample was a high-functioning, healthy, and highly educated group of older adults, and the same tasks were used at each wave, allowing for practice effects to occur.

The potential influence of the short time period, select sample, and possible practice effects were even apparent in the lack of significant age group differences in cognitive change. Although the old-old age group (i.e., 75-92 years) achieved lower cognitive scores on all five cognitive tasks at Wave 1, age group effects were only found on the word recall and vocabulary tasks. The average young-old participant (i.e., 64-74 years) recalled more words and correctly identified more word definitions with each year of being in the study, while the average old-old participant showed a smaller gain in word

recall, and a slight decline in vocabulary performance. However, research indicates nearly all cognitive domains show accelerated decline in old-old adulthood (e.g., Salthouse, 2004). Overall, the modest and selective age group effects reinforced the selectiveness of the present sample and the possibility that significant age-related cognitive change tends to occur over longer time intervals.

Because we used hierarchical linear modeling (HLM), we could also evaluate whether significant change occurred at the individual level (i.e., did individuals show different slopes in cognitive performance over time?). For all five cognitive domains, there were significant individual differences in both the initial level and change in cognitive ability over 3 years. Clearly, not all participants followed the same average slope (e.g., see Figure 1). Some individuals did decline as expected over the 3 years, but others improved or maintained their cognitive performance. The presence of both between-individual differences in change and significant average effects indicated 3-year cognitive change did in fact occur. These change effects, together with significant within-person error (i.e., fluctuations from each individual's best-fitting regression line) that remained to be explained after accounting for time in study, suggested significant variance remained to be explained by other predictors (e.g., intraindividual variability).

Inconsistency

Despite expectations for significant overall change in intraindividual variability, the motor, basic, and complex composites were stable over the 3 years. Inconsistency on the interference tasks did show a small positive increase. The scarcity of significant average increases in intraindividual variability however can be reconciled with previous longitudinal research. MacDonald and colleagues (2003) found inconsistency increased across 6 years only for those aged 75-89 years at baseline, but remained stable or

decreased slightly for those 55-64 years and 65-74 years. Similarly, Lövdén and colleagues (2007) found intraindividual variability steadily increased across 13 years in those aged 70-102 years. The present sample was aged 64-92 years, with just under half of the participants aged 75 or older. Consequently, the failure to find average age effects may be due the younger average age of the sample (i.e., 74.29 years). In fact, we did find significant age group differences in both the baseline level and change in inconsistency over the 3 years. Old-old adults (i.e., 75-92 years) showed higher initial levels of intraindividual variability on all four composites than the young-old adults (i.e., 64-74 years), consistent with cross-sectional research (e.g., Hultsch et al., 2002). Further, although there were no age group differences in change on the motor and interference ISDs, the average old-old participant showed significantly greater increases in inconsistency on the basic and complex composites. In comparison to the average old-old participant, who experienced modest increases in variability in responding with each additional year of being in the study, the average young-old participant showed no change on basic ISD, and actually experienced a slight decrease on complex ISD. The fact that age group differences were most apparent on higher complexity RT tasks is also in accordance with MacDonald et al., who found 6-year change in inconsistency was greater on the more complex verbal tasks (i.e., lexical and semantic decision) rather than the non-verbal tasks (i.e., simple and choice RT).

Overall, the consistent evidence from both the present findings and the two existing longitudinal studies on inconsistency (Lövdén et al., 2007; MacDonald et al., 2003) clearly suggest that significant group-based increases in intraindividual variability are limited to the latter half of older adulthood (i.e., age 75 and older). Therefore, intraindividual variability does not uniformly increase across older adulthood, but rather

accelerates as one ages into old-old adulthood. The present study also demonstrated that these age-related changes can be observed across even as short a time period as 3 years. However, because these increases were only apparent on cognitively challenging RT tasks, the underlying neural causes of inconsistency likely only experienced modest age-related changes.

Of course, not all participants had to experience the same rate of change in inconsistency over time. There were also significant random effects, or between-person differences in the starting value and slope of inconsistency performance for all four ISD composites. That is, individuals significantly differed in their initial amount of variability, and also whether they increased, decreased, or maintained their level of intraindividual variability over the 3 years (e.g., see Figure 2). Therefore, even if significant change was not evident at the group level (i.e., for motor, basic, and complex ISD), significant individual differences indicated there was sufficient variation in change to next investigate possible covariation with cognition over the 3 years.

Covariation of Cognition and Inconsistency

If short-term fluctuations in performance truly are markers of long-term cognitive change, the two measures must be significantly linked, such that changes in one appear to correspond to changes in the other. Demonstrations of the significant coupling relationship have been found across 2- (Lövdén et al., 2007) and 3-year intervals (MacDonald et al., 2003), such that regardless of the time point, increases in inconsistency across the time intervals were related to decreases in cognitive ability. The present study extended this research by investigating a shorter testing interval (i.e., annual change), thus further exploring the sensitivity between intraindividual variability change and short-term cognitive change.

Using HLM, we found the overwhelming majority of the models (i.e., 17 out of 20) showed significant covariation between change in cognition and change in intraindividual variability. In each of these cases, the slope estimates were negative, in that for each additional unit increase in intraindividual variability (i.e., 0.1 SD) participants' cognitive performance correspondingly declined. Further, the negative covarying relationship between cognitive performance and ISD was invariant across time; regardless of the time in study, at testing occasions when a participant's variability increased, their cognitive performance also decreased.

These results replicate the significant coupling found by Lövdén et al. (2007) and MacDonald et al. (2003) that inconsistency is indeed related to cognitive ability over time, and that the two are negatively associated. The coupling relationship suggests that the two measures may be caused or influenced by the same underlying process, which is likely neurological in nature based on brain-based research (e.g., Anstey et al., 2007; Bellgrove et al., 2004). However, the present results do not rule out the presence of two separate causes or that the two processes cause one another. The present study extended the time frame of previous research, demonstrating that the link between inconsistency and cognition is stable across even as short a time frame as one year. The short temporal coupling between the two measures suggests that intraindividual variability may not only be an early marker of cognitive decline, but a particularly valuable one, where changes in one are reliably reflected in the other in as short as yearly measurements. Future research could investigate the minimum time interval for significant covariation between the two measures. If both inconsistency and cognition are believed to be driven by neurological mechanisms, it is not unreasonable to hypothesize that the two will demonstrate coupling even across the same day. Finally, although the present study cannot address the lead-lag

relations between inconsistency and cognition, Lövdén and colleagues found increased inconsistency reliably preceded cognitive decline, rather than vice versa, indicating changes in inconsistency may be key indicators of later cognitive decline in older adulthood.

The coupling relationship was surprisingly stable across the older adult age range, as only 2 of the 20 models had significant age group effects (i.e., motor ISD with letter series, interference ISD with vocabulary). Because the brain undergoes dedifferentiation, or compression of the organization of cognitive abilities in older adulthood (Li, Lindenberger et al., 2004), it was reasonable to expect the covariation relationship might similarly grow stronger at that time. However, consistent with MacDonald et al. (2003) and Lövdén et al. (2007), the young-old adults experienced the same unit decrease in their cognitive score with each unit increase in inconsistency as the old-old adults, indicating the relationship between cognition and intraindividual variability remained stable across the older age range. This age-invariant relationship is even consistent with substantially greater increases in inconsistency in old-old adulthood (e.g., MacDonald et al., 2003) because there are also corresponding larger declines in cognitive ability past 75 years of age. The stability of the coupling relationship from age 64-92 is interesting, because it suggests similar covariation may exist across the entire age range. In addition, it remains to be seen whether the link between variability and cognition remains stable even in pathological populations. Initial impressions from the well-substantiated link between greater inconsistency and poorer cognitive performance in a variety of age ranges (e.g., Hultsch et al., 2002) and clinical conditions (e.g., dementia, Hultsch et al., 2000; Parkinson's, Burton et al., 2006) indicate these ideas are promising. Further there appears

to already be convincing evidence as the results presented by Lövdén and colleagues were not confounded by suspected dementia or time-to-death.

Differences Due to Task Complexity

The annual coupling relationship was also quite reliable, as it was evident across the continuum of cognitive tasks and inconsistency measures. However, the magnitude of the relationship varied according to the cognitive complexity of the paired measures. Interestingly, this influence was apparent in the difficulty of both the cognitive tasks and the inconsistency-based tasks.

Impressively, there were significant coupling relationships found across all cognitive tasks. Digit symbol showed the strongest resulting change in slope for a one unit increase in ISD, while vocabulary showed the smallest corresponding changes. In terms of the strongest cognitive domain, MacDonald et al. found memory-based measures, such as word recall, showed the strongest 3-year coupling with intraindividual variability, followed by working memory, perceptual speed, and reasoning. Rather, for the present study, perceptual speed (i.e., digit symbol) showed the strongest effects, followed by episodic memory and reasoning. Therefore, it is still unclear as to which cognitive domain shows the strongest relationship with intraindividual variability. It is clear however, which domain has the weakest relationship. MacDonald and colleagues (2003) found the 3-year covariation between inconsistency and cognition was significant across a range of cognitive domains, but they did not find a significant association with crystallized verbal ability (vocabulary). On the other hand, Lövdén et al. (2007) found significant effects for both of the assessed cognitive domains, one of which tapped crystallized knowledge (i.e., ideational fluency). It may be the case that the coupling relationship between inconsistency and vocabulary type tasks is limited to 1- and 2-year

time frames because of its small effect. Vocabulary showed the smallest corresponding 1-year changes in the present study, and similarly showed modest changes compared to that for perceptual speed in 2-year covariation (Lövdén et al., 2007). It appears that domains that assess fluid abilities such as memory, reasoning, and processing speed are more tightly linked to inconsistency in performance, just as they show more age-related decline.

For inconsistency, fluctuations on the basic and complex RT tasks demonstrated the strongest coupling relationships across the cognitive domains. Pairwise comparisons of the fixed ISD slope estimates showed complex ISD and basic ISD consistently outperformed motor ISD, which was calculated from the simplest RT tasks (i.e., motor tapping). Compared to interference ISD, both complex and basic ISD were stronger on word recall and digit symbol, but only complex ISD showed a stronger relationship than interference ISD on the letter series, similarities, and vocabulary tasks. Basic ISD did not significantly differ from interference ISD on these tasks. Overall, the moderately complex tasks that formed basic ISD (i.e., variations of CRT) and the highly complex tasks that formed complex ISD (i.e., required inhibiting a previous or current response) both provided an equally good index of change in the cognitive domains of episodic memory and processing speed. For the other cognitive domains of reasoning, verbal fluency, and vocabulary, inconsistency from the most challenging RT tasks (i.e., complex ISD) showed the strongest coupling of change.

It was interesting that interference ISD, which is based on fluctuations in responding to the arrow stroop task, was a poorer measure of coupling cognitive change than the basic and complex ISDs. Even more surprising was that interference ISD did not significantly differ from the simplest ISD composite (i.e. motor ISD) for four of the five

cognitive tasks. The arrow task required inhibiting the spatial composition of the stimulus, and responding only to its figural form (i.e., the direction of the arrow). The present results clearly show that the cognitive demands from this type of inhibition were minimal, and surprisingly far less than those required to perform a choice RT task, and barely more than those required to complete a simple RT task. These results are different from those reported by Strauss et al. (2007), who used the same four ISD composites, but found the complex ISD, followed by the interference ISD, provided the best differentiation of various subtypes of potential preclinical forms of dementia. However, recent analyses by our research group using the arrow stroop task have failed to find the anticipated stroop effect (i.e., slowest responding to the incongruent condition), indicating the task itself may be flawed.

Overall, it is clear that complexity plays a role in the coupling relationship between change in intraindividual variability and change in cognition. These results are consistent with findings that inconsistency based on cognitively demanding tasks provided greater sensitivity to cross-sectional age differences (e.g., Bunce et al., 2004; West et al., 2002), and the best discrimination among varying degrees of mild cognitive impairment (Dixon et al., 2007; Strauss et al., 2007). There are caveats to this conclusion for the covariation relationship however. The complexity of the ISD-based tasks appears to interact with the complexity of the cognitive domain. The impact of the complexity of the inconsistency measure was most evident on cognitive domains that did not decline with age (e.g., crystallized abilities) and reasoning. For the other cognitive domains (i.e., episodic memory, processing speed), which were arguably those that tend to show the greatest decline with older age, inconsistency on the moderately complex RT tasks were similar to inconsistency on the more challenging RT tasks. Variability in responding

while under high cognitive demand may be most attuned to the integrity of the neurological system, and this type of sensitivity appears to be accordingly required to show significant coupling with those cognitive domains that show relatively subtler changes with older age.

Limitations and Future Directions

In terms of limitations, the present sample was relatively well-educated and healthy compared to the general population. The selectiveness of the sample was further demonstrated by observing significant improvements in cognitive performance, rather than decline, and the adults showing only modest changes in intraindividual variability. Given this limited change, one might expect a lower likelihood of observing significant effects and coupling between change in inconsistency and change in cognition. However, finding significant results even with such a selective sample bolsters the present results, and suggests replication attempts in more diverse and less healthy samples will potentially show even stronger effects. This would prove a fruitful area of future research.

As mentioned earlier, hierarchical linear models cannot address lead-lag relationships, or whether changes in one variable might have preceded changes in the other. Therefore, the present results do not provide information about the direction of the change relationships. However, recent findings using bivariate dual change score models, which do allow for corresponding time-lagged associations between the variables, show changes in inconsistency appear to precede changes in cognitive ability (Lövdén et al., 2007). Replication of this temporal relationship is needed.

Overall, the present study substantially enhanced the knowledge surrounding intraindividual variability and older adulthood. There were significant individual

differences in short-term 3-year change in inconsistency, but group-based increases were limited to the latter half of older adulthood (i.e., approximately 75 years and older). There was further verification of the longitudinal link between inconsistency and cognitive ability, demonstrating that the two measures significantly covaried across intervals as short as one year. Further, this coupling relationship remained stable across mid-old to old-old adulthood, suggesting similar stability may exist across the entire age range, and even pathological populations. The strength of the coupling relationship does appear to differ though based on the cognitive difficulty of both the cognitive domain and the inconsistency-based tasks. Future research could investigate how stable the coupling relationship is over shorter time intervals, the lead-lag relationship between cognition and intraindividual variability, and extending this research to other age groups and populations.

Study 2

The first study demonstrated that annual change in short-term intraindividual variability was indeed related to annual intraindividual change across a range of cognitive abilities. Of particular interest was the short time interval separating the testing occasions, suggesting that intraindividual variability is highly sensitive to even subtle changes in cognitive ability. Given the sensitive nature of the coupling relationship, is intraindividual variability also sensitive to change in other meaningful outcomes, such as incipient disease? There is substantial evidence that increased inconsistency is associated with maladaptive traits and health conditions, including poorer physical performance (e.g., Li et al., 2001), less lifestyle engagement (e.g., Bielak, Hughes et al., 2007), traumatic brain injury (e.g., Stuss et al., 2003), dementia (e.g., Hultsch et al., 2000), Parkinson's disease (e.g., Burton et al., 2006), mild cognitive impairment (e.g., Dixon et al., 2007), and inferior structural brain characteristics (e.g., Anstey et al., 2007), but these associations have all been based on cross-sectional research. Many have echoed Nesselroade and Salthouse's (2004) claim that the literature "needs to move towards building predictive relationships" (e.g., Hultsch & MacDonald, 2004; Martin & Hofer, 2004). In fact, only one study has investigated the predictive link over time. Using survival analysis, MacDonald and colleagues (in press) found baseline intraindividual variability in cognitive performance could predict impending death up to 15 years later. Together with these impressive results, the significant cross-sectional associations with various conditions, and verification that inconsistency precedes cognitive decline (Lövdén et al., 2007), the potential for inconsistency to predict other meaningful outcomes is promising.

There has also been much discussion in the intraindividual variability literature about whether level of performance (i.e., mean RT) provides the same information as inconsistency. Since the measures are typically highly related to one another, many have argued that short-term variability may not offer any unique predictive power beyond mean level of performance (e.g., Salthouse et al., 2006). Consequently, studies have often evaluated inconsistency's relationship to various outcomes compared to that with the mean, or over and above mean level. Although many studies have found inconsistency does predict outcome measures independent of mean level (e.g., Burton et al., in press; Hultsch et al., 2002; Li et al., 2001), some studies have found negligible differences between the two measures (e.g., Bielak, Hughes et al., 2007; Christensen et al., 2005). However, in possibly the strongest investigation to date, Lövdén et al. (2007) found higher inconsistency reliably preceded and predicted subsequent decline in cognition, but mean performance did not significantly predict later cognitive decline. Therefore, inconsistency may be stronger at reliably predicting meaningful change outcomes.

The second study investigated which types of meaningful 5-year change outcomes could be predicted by baseline intraindividual variability, and compared these results with those using mean level of performance. We were specifically interested in whether inconsistency could predict later cognitive status, and possibly serve as an early indicator of cognitive decline, mild cognitive impairment, or detrimental health outcomes. A range of classification schemes for cognitive status were evaluated including the direction of change in each domain of cognitive performance (i.e., improve, stable, or decline), change on the Mini-Mental State Examination (i.e., improve, stable, or decline), and change in cognitive status classification (i.e., stable intact, fluctuate, stable decline, stable

CIND). The MMSE is frequently employed in medical settings (e.g., doctor's offices) to obtain a quick measurement of overall cognitive functioning. Therefore, a potential link between MMSE change and inconsistency would be quite valuable and pragmatic. Similarly, research has shown inconsistency can distinguish between individuals with various subtypes of MCI status and healthy older adults (Christensen et al., 2005; Dixon et al., 2007, Strauss et al., 2007), but the stability of the MCI classification from year to year is poor. Given the possibility that intraindividual variability is a sensitive marker of neurological integrity, it may be also be attuned to poorer yet reliable change patterns over 5 years, thus identifying individuals with valid MCI classifications.

We also wanted to investigate whether intraindividual variability could predict later dementia diagnosis, but few individuals in the present sample transitioned to a formal diagnosis over the 5 years. Therefore, we instead used attrition as a proxy for impending decline. Sliwinski et al. (2003) noted that because deleterious events and pathological changes are more likely to occur in older adulthood, they increase the likelihood of individuals dropping out of longitudinal studies. Consequently, attrition may be an indicator of less easily identifiable nonnormative influences on cognitive decline in older adulthood, and tends to reflect underlying influences such as disease and cognitive impairment (MacDonald et al., 2003). In fact, Sliwinski et al. (2003) found attrition-based cognitive change accelerated 4-6 years before drop-out and became an increasingly influential predictor of change; more so than chronological age and time to death. Therefore, just as attrition appears to be a sensitive indicator of impending problems, intraindividual variability may be predictive of attrition.

The predictive power of intraindividual variability was also contrasted with that from the corresponding mean RT scores at Wave 1, and we investigated whether

cognitively challenging tasks were superior predictors of impending cognitive change. Finally, the present study addressed whether inconsistency offered better prediction of attrition than standard neuropsychological tests. Although studies investigating inconsistency are becoming more apparent in the clinical literature (e.g., Christensen et al., 2005; Murtha et al., 2002), most clinicians do not administer tests that can calculate short-term, trial-to-trial inconsistency, such as RT tasks. Moreover, completing these additional tasks can add unnecessary burden on the client or patient, and the calculation of inconsistency can be time-consuming. To date, there are no studies that provide concrete evidence that clinicians should include or substitute intraindividual variability measures as part of their diagnostic inventory. This final question will address whether intraindividual variability is more attuned to possible nonnormative age-related changes (i.e., attrition) than standardized and widely used neuropsychological indicators.

Research Questions

1. Does inconsistency at Wave 1 predict the following cognitive outcomes 5 years later?

a) Level of cognitive performance at Wave 6

This question aimed to replicate the initial findings of study 1 by extending them to 5 years. Based on those findings and significant relationships found over 6 years (MacDonald et al., 2003), we expected inconsistency at Wave 1 to significantly predict level of cognitive ability at Wave 6.

b) Change in cognitive performance

Building on the positive change relationships found between cognitive performance and inconsistency over 6 years (MacDonald et al., 2003), we hypothesized that inconsistency would be similarly effective at predicting reliable cognitive change

group over time. Specifically, we expected Wave 1 inconsistency to significantly differentiate among individuals who declined, maintained the same level of cognitive performance, or improved across the 5 years, where greater inconsistency was associated with a poorer change group. Further, because the vocabulary domain showed the weakest covarying relationship with inconsistency in study 1 (consistent with MacDonald et al., 2003), we expected greater predictive relationships would exist for change in fluid processing measures (i.e., episodic memory, reasoning, perceptual speed) compared to crystallized measures like vocabulary, and to a smaller extent, verbal fluency.

c) MMSE change

Previous studies have shown intraindividual variability is very good at predicting even subtle changes in pathological cognitive decline (e.g., CIND, Strauss et al., 2007), so similar associations were expected with change on general cognitive measures like the MMSE. It was expected that inconsistency would be able to distinguish between groups who show maintenance of their ability over time, clinically significant declines in performance, and improvement.

d) Cognitive status change

The MCI classification remains controversial as some MCI individuals fluctuate back to normal classification rather than continuing to show symptoms consistent with MCI. Therefore, which individuals truly have MCI is unknown. However, given that individuals with poorer cognitive status (i.e., CIND status with multiple domains of impairment) tended to be more variable on cognitive tasks (e.g., Dixon et al., 2007), we similarly expected inconsistency to significantly distinguish between those with adaptive and maladaptive classification patterns. That is, we expected inconsistency at Wave 1 to distinguish between individuals who remain classified as CIND, those who become CIND

over time, those who fluctuated in status classification, and those who remained with an intact cognitive classification over the 5 years.[5] Specifically, we expected those with maladaptive 5-year change patterns to show greater intraindividual variability in responding.

e) Attrition

MacDonald and colleagues (2003) found individuals who later dropped out of the 6-year longitudinal study were more inconsistent at the baseline assessment. Therefore, similar relationships were expected in the present sample.

2. Is inconsistency a better predictor of the outcomes [i.e., a), b), c), d), and e)] than mean level of performance?

The recent finding that inconsistency, but not mean level, reliably preceded and predicted subsequent decline in cognition (Lövdén et al., 2007) led us to expect inconsistency would similarly be a better predictor of the group outcomes than mean level of cognitive performance.

3. Does the level of cognitive effort of the tasks intraindividual variability and mean level are based on result in different strength in predicting the cognitive group outcomes?

Strauss et al. (2007) showed stronger cross-sectional relationships of inconsistency with CIND subtypes for tasks with greater cognitive complexity. Similarly, study 1 found the strongest coupling relationship with cognition existed for the inconsistency measures based on moderately complex RT tasks. Therefore, we expected

[5] In light of the lack of consensus surrounding preclinical cognitive impairment, we have chosen to use the term Cognitive Impairment, No Dementia (CIND), which is a more general, inclusive term that encompasses many of the various definitions (Tuokko & Frerichs, 2000), including Mild Cognitive Impairment (MCI; Petersen et al., 1999).

these same tasks (i.e., basic and complex composites) to also have better predictive ability than the simpler RT tasks.

4. Does inconsistency predict attrition better than neuropsychological tests?

This research question has not been evaluated before, even in a cross-sectional format, so the predictive power of inconsistency compared to neuropsychological tests remained to be seen. However, given the strong relationship between intraindividual variability and later cognitive capacity, inconsistency at Wave 1 was expected to predict group membership quite comparably to Wave 1 neuropsychological test scores.

Method

Study 2 used the same sample, and similar measures and procedures as described in Study 1. However, Study 2 expanded the time frame and used data collected at the 6th wave of measurement (i.e., 5 years later). Additional information about the methodology for Study 2 is described.

Participants[6]

212 participants completed all of the relevant Wave 6 tests[7]. At this time point, the majority of participants were now in the old-old age group classification: Young-old (65-74) $n = 65$, $M = 72.38$, $SD = 1.35$; Old-old (75+) $n = 147$, $M = 81.31$, $SD = 4.67$. As expected, women made up the majority of the sample (69.8%). The participants were well educated ($M = 15.26$, $SD = 3.02$), ranging from 7 to 24 years of education, with only 9.4% having less than 12 years of schooling ($n = 20$). Nearly half of the sample was married or common-law (51.9%), 29.2% were widowed, 16.1% were divorced or

[6] Group outcome e) attrition used the entire sample of 304 participants.
[7] 6 additional participants completed various components of the wave 6 test battery, but did not complete all relevant measures for the present study. To maintain consistency across the analyses, these participants were excluded.

separated, and 2.8% were single. The participants were relatively healthy, with few reported chronic disorders (M =3.22, SD = 1.94), and 60.8% having three or fewer medical conditions. The participants' perceptions of their relative health were also high, with 80.7% rating themselves good or very good compared to a perfect state of health, and 94.3% rating themselves good or very good compared to the health of other's their own age. Finally, most participants reported they were very capable of completing instrumental activities of daily living (Lawton & Brody, 1969); more than 88% were able to shop, prepare food, complete laundry and housekeeping, oversee their own medications and finances, and drive or travel independently.

Procedure

The fifth and sixth waves of measurement included relatively fewer tasks and sessions. In the fifth testing wave, participants only completed a brief telephone interview. Due to the scarcity of tasks completed at the fifth wave, data from this wave was not used for the present study. The next occasion however, Wave 6, involved two group testing sessions. This testing session included re-administration of the five cognitive tasks already tested in previous waves, a general cognitive functioning questionnaire, and a personal health update, along with other brief questionnaires and tests. The majority of these sessions were administered at the university, but at the participant's request, approximately one third of the sample was administered the test battery individually in their homes.

Measures

The same 5 cognitive tasks (i.e., those assessing perceptual speed, reasoning, episodic memory, verbal fluency, and vocabulary), and all RT tasks described in the first

study were used. Other measures relevant to the construction of the outcome groups and the comparison neuropsychological tests are described.

Because very few participants were diagnosed with dementia by Wave 6, attrition was used as a potential indicator of cognitive impairment and impending dementia. The reasons for participants' drop-out at each wave could include a range of other possible causes including loss of interest, illness, and increased demands on their time. However, to test for the current possibility that dementia could have been one of the possible causes for participant's attrition, dementia screening tests were used. The clinical criteria for dementia (American Psychiatric Association, 1994) involve the presence of multiple cognitive deficits, including memory impairment. Research investigating the best neuropsychological predictors similarly recommends a wide range of indicators. Marcos et al. (2006) found significant differences in the initial MMSE, memory, attention, and abstract thinking assessments of MCI individuals who went on to develop dementia 14.6 months later, and those individuals who remained as stable MCI. Bäckman, Jones, Berger, Laukka, & Small's (2005) list of predictive cognitive domains went even further, as their meta-analysis verified that there are marked preclinical deficits in global cognitive ability, episodic memory, perceptual speed, executive functioning, and smaller deficits in verbal ability, visuospatial skill, and attention. Further, researchers have identified specific neuropsychological tests that are sensitive to early AD changes, such as finding that mild AD patients had normal performance on Part A of the Trail Making Test, but marked impairments compared to normals on Part B of the task (Baudic et al., 2006).

Based on findings from this body of literature, the following neuropsychological indicators were employed:

Mini-Mental State Examination (MMSE; Folstein et al., 1975). This task is a general indicator of overall cognitive functioning, and is primarily used to detect gross cognitive impairments. It assesses a range of cognitive skills, including questions pertaining to orientation, short-term recall, working memory, and spatial ability. The total possible score was 30.

Word recall. The word recall task described in Study 1 (Hultsch et al., 1990) assessing episodic memory was used.

Trail Making Test - Part B (TMT-B; Reitan, 1958). This task is an indicator of executive functioning, and requires inhibition and other cognitive demands such as visual-spatial ability. Participants were asked to draw lines linking numerical and alphabetical circles on the test paper, but alternate between numerical and alphabetical sequencing (e.g., 1, A, 2, B). The outcome measure was the time taken to complete the task.

Wechsler Adult Intelligence Scale-III (WAIS-III) Block Design (Psychological Corporation, 1997). This subscale assesses visual construction and visuospatial skill. Participants were given coloured blocks and asked to replicate designs shown in a booklet.

WAIS-III Vocabulary (Psychological Corporation, 1997). This subscale is an indicator of verbal ability. Participants were asked to provide the definitions of given words. Total correct was the outcome measure.

Construction of Outcome Groups

Outcome variables a) *level of cognitive performance*, and b) *change in cognitive performance*, were derived from the overall level of performance on each of the 5 cognitive tasks administered at a) Wave 6, and b) Waves 1 and 6, respectively. Level of

cognitive performance used the numerical value achieved by each participant on each of the 5 tasks. Change groups in cognitive performance on each of the 5 cognitive tasks were determined by first calculating whether the observed change was reliable enough to be considered significant beyond measurement error (e.g., practice effects). In the aging literature, change has been considered reliable if it exceeds 1.0 standard error (*SE*) of measurement for the sample (e.g., Schaie, 1989). On the other hand, the clinical literature uses a well known variation of this technique, the reliable changes index (Jacobson & Truax, 1991), which focuses on the standard error of the difference score, which then is multiplied by 1.96 to obtain a 95% confidence interval. However, Hawley (1995) argued that the reliable change index is not applicable when working with a nonclinical population. Reliable change indices originally assumed movement from a dysfunctional distribution to one that is functional (e.g., improvement via therapy), and thus required a dramatic change to be deemed reliable (i.e., 2 SDs). Consequently using the same criteria for establishing a threshold of change in the present normal population was inappropriate. In fact, the cutoff values for each cognitive task produced from 1.0 SE of measurement (aging literature) or 1.0 SE of the difference score (a variation of the reliable change index) were nearly identical; the values calculated from 1.0 SE of measurement tended to be .15 smaller than those calculated using 1.0 SE of the difference. Further, given that the 5 cognitive tests were evaluated in whole numbers, the difference in the threshold for reliable change did not significantly impact the consequent groupings. As a result, 1.0 SE of the difference was used to determine reliable change on each 5 cognitive tasks.

Separate calculations were conducted for each task, by each Age X Education group (age groups = 65-74 years and 75+ years; education groups = 0-12 years and 13+ years). We expected old-old adults and those with less education to show greater decline

over time, and thus correspondingly chose to use reliable change values that reflected these expectations. Further, we did not want to overlook modest but reliable changes in the young-old and better educated by using a larger SE of the difference that was based on the entire sample[8].

The following equations were used

SE: $SD_1 \sqrt{(1-r_{12})}$

$SE_{difference} = \sqrt{2}(SE^2)$

where SD_1 is the standard deviation across all participants at time 1, and r_{12} is the Pearson's bivariate correlation across all participants between test scores at time 1 and time 2. The $SE_{difference}$ was then calculated.

Individual change in cognitive score for each task was calculated (Wave 6 – Wave 1). These numbers were compared to the resulting cutoff values, and participants were classified into the following 3 groups for each of the five cognitive tasks (see Table 6 for the cell sizes for each cognitive task):

1) Stable – performance change within $\pm 1.0\ SE_{difference}$. The change that has occurred is not significant.

2) Improve – performance change greater than $+1.0\ SE_{difference}$. There has been a significant improvement in performance over time.

3) Decline – performance change greater than $-1.0\ SE_{difference}$. There has been a significant decline in performance over time.

[8] We compared the group classifications according to the overall SE threshold values with those determined by cutoffs for Age X Education group. The distinction among groups was slightly enhanced by using thresholds created for each Age X Education group, as more individuals moved from the stable grouping to showing either reliable improvement or decline.

Table 6

Cell Sizes for each Change Status per Cognitive Outcome Change Group

	Change status		
	Improve	Stable	Decline
Cognitive Measure	*n*	*n*	*n*
Digit Symbol	61	121	30
Letter Series	30	146	36
Word Recall	43	135	34
Similarities	39	135	38
Vocabulary	38	142	32

Outcome variable c) *MMSE change* was based on MMSE scores at Waves 1 and 6. Individuals were divided into one of three groups based on the amount of change in their MMSE score across the 5 years. In accordance with other research (e.g., Clark et al., 1999), significant clinical change was defined as a minimum decline of 3 points.

1) Stable – zero change or 2 point decline in performance over time (n = 151).

2) Improve – 1 or more point increase over time (n = 44).

3) Decline – 3 or more points decline in performance over time (n = 17).

Outcome variable d) *Cognitive status change* was based on potential variations in the participants' cognitive status classification over the 5 years. First, cognitive status at each testing wave was determined by participants' performance on the 5 cognitive tasks (WAIS-R Digit Symbol Substitution, Wechsler, 1981; Letter Series, Thurstone, 1962; immediate free word recall, Hultsch et al., 1990; Controlled Associations, Ekstrom et al., 1976; and Extended Range Vocabulary, Ekstrom et al., 1976). Their performance was compared to norms obtained from an independent sample of 445 adults aged 65 to 94 years drawn from the same population[9]. Although published norms are available for most of these measures, they are derived from a variety of different samples with varying demographic characteristics. The use of local norms from a separate sample drawn from the same population is preferred given the close demographic match to the current sample and the ability to make more accurate comparisons across tasks. The normative sample was partitioned into four Age x Education groups (age groups = 65-74 years and 75+ years; education groups = 0-12 years and 13+ years). Means and standard deviations

[9] Data on all 445 participants of the normative sample were available for the measures of perceptual speed, inductive reasoning, verbal fluency, and vocabulary, but due the use of a counterbalancing procedure, information on the episodic memory task was only available for 194 of the 445 participants. Individuals participating in the longitudinal study from which the normative sample was drawn were not accepted into the current sample.

were computed for these groups for the five measures and these normative values were used to classify participants from the present sample on the basis of cognitive status.

There were two ways an individual could be classified as possible CIND. First, participants whose performance was more than 1.5SD below the mean of their age- and education-matched peers on only one cognitive reference task fit the criteria for CIND-Single. The 1.5SD criterion has been widely used in the clinical literature (e.g., Peterson et al., 1999; Tuokko, Gabriel, & the CSHA Neuropsychology Working Group, 2006) and represents a stricter criterion than those previously used with this dataset (e.g., Bielak, Hultsch, Kadlec, & Strauss, 2007). Further, previous classification using only the 1.0SD criterion in a single cognitive domain has resulted in few significant differences in cognitive performance between the CIND-Single individuals and those with no cognitive impairments (e.g., Bielak, Hultsch et al., 2007). A stricter criterion was hypothesized to solve this difficulty. Second, participants who scored more than 1.0SD below the normative sample on two or more cognitive tasks fit the criteria for CIND-Multiple. The CIND-Multiple cut-off remained as 1.0SD as previous studies with this dataset have demonstrated significant differences in cognition do exist between the CIND-Multiple group and normal older adults (e.g., Strauss et al., 2007). Further, the 1.0SD CIND-Multiple guideline has been used alongside the 1.5SD CIND-Single classification in population-based studies (e.g., Canadian Study of Health and Aging [CSHA]). All remaining participants who did not fit the criteria for either CIND group were classified as cognitively intact (Intact).

Next, for the purposes of detecting change in CIND classification over time, the CIND-Single and CIND-Multiple groups were collapsed into one CIND group. This procedure follows that employed by the Canadian Longitudinal Study of Health and

Aging (see Tuokko et al., 2006). Further, given that it was possible to be in both the CIND-Single and CIND-Multiple groups at the same wave of testing (i.e., a participant could show greater than 1.5SD impairments in only one cognitive domain, and also show greater than 1.0SD impairments in two or more cognitive domains), combining the two CIND classifications was necessary to classify participants into CIND change groups.

The pattern of change of the CIND classification over the 5-year period was assessed using classifications from Waves 1, 3, 4, and 6[10]. Individuals were classified according to the following potential patterns of change (see Figure 4):

1) Stable intact – the participant was consistently classified as intact (n = 118).

2) Fluctuating – the participant's cognitive status classification fluctuated from wave to wave, fluctuating from CIND to normal and vice versa (n = 58).

3) Stable decline – the participant was initially classified as Intact and eventually became CIND over time (n = 15).

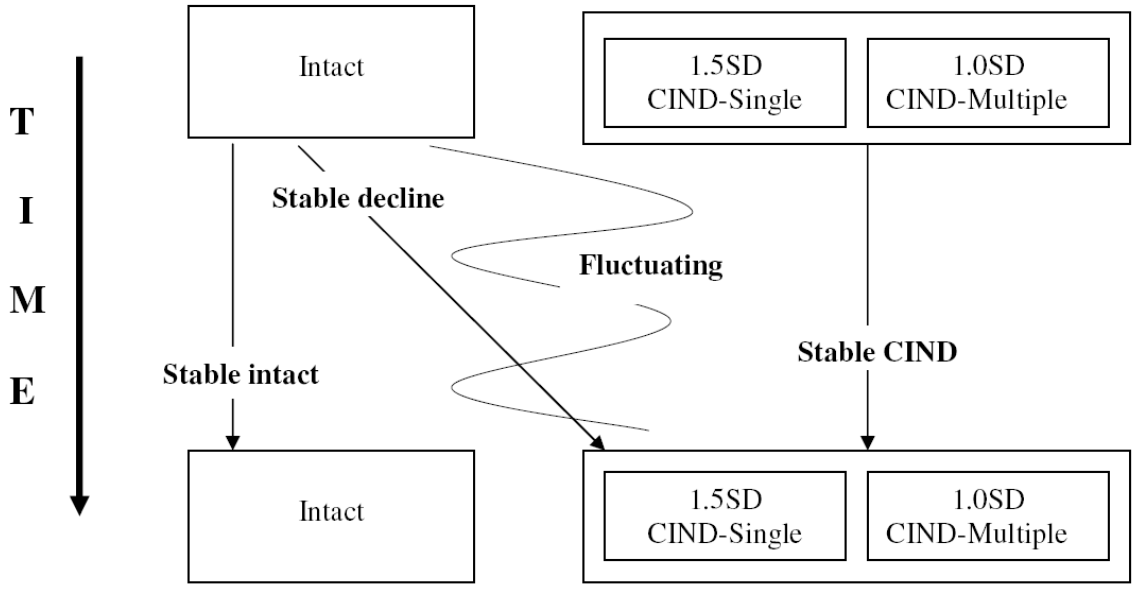
4) Stable CIND – the participant was consistently classified as CIND (n = 21).

Because the current classification of cognitive impairment is imprecise and thus prone to fluctuation, at least 4 classification patterns warranted comparison. First, individuals who consistently showed average cognitive ability, and those who consistently showed poor cognitive ability relative to their peers likely represent two distinct groups. Next, it was important to distinguish between the two possibilities of fluctuating classification patterns between these two extremes: individuals whose

[10] Due to budgetary difficulties, 10 participants were not retested in Wave 2, but continued on in Waves 3, 4, and 6. Rather than exclude these participants from the present study, we chose to exclude Wave 2 CIND classifications from the CIND change pattern. The remaining 4 years were deemed to be sufficient to observe change in CIND status (i.e., Waves 1, 3, 4, and 6), and given that both the wave preceding and immediately following Wave 2 were included, any changes in CIND status would be detected.

Figure 4

Possible Change Trajectories in CIND Classification over 5 Years



classifications changed dramatically across the waves may represent individuals on their way to cognitive impairment (i.e., group 2 - Fluctuating), but their outcome may be less certain than those who became CIND over time and continued to be classified this way (i.e., group 3 – Stable decline).

The final outcome variable e) *Attrition* was based on whether participants remained in the study and completed Wave 6 testing. There were 86 participants who did not complete the test battery given at Wave 6, leaving 218 continuing participants. Not all of the continuing participants completed all 6 measurement occasions, but they reliably attended the study across the 5 years. For the non-continuing participants, their date of attrition was determined one of two ways, according to their last testing date: 1) If the participant completed a testing wave (i.e., group, intact, and individual sessions), but did not return for the next wave of testing, their date of attrition was calculated as one year after their last testing date; or 2) If the participant stopped attending the testing sessions midway through the test battery (e.g., attended the group, intake, and their first individual sessions, but did not complete the rest of the individual sessions), their date of attrition was calculated as 2 weeks after their last testing date[11].

The date of attrition was skewed however for ten of the non-continuing participants. Because the fifth wave of testing involved only a telephone interview, which required little time and energy, there were ten participants who completed that interview, but had not returned to study testing for some time. For example, there were 6 participants who completed only Waves 1 and 2, and then completed the phone interview 3 years later. Given that the attrition analysis was designed to identify individuals who may be at risk for cognitive impairment (i.e., cognitive difficulties may be a reason

[11] The lag time between testing sessions was approximately 2 weeks.

participants refused to return for testing), it was decided that completing the telephone interview did not qualify as a sufficient return to testing for these ten participants. Furthermore, if we accepted the telephone interview as a sufficient return to testing, the participant's date of attrition would have been delayed by 2 to 5 years, substantially skewing their date of attrition.

Data Preparation

Only Wave 1 RT data was used in the present study. The data trimming procedures were identical to those in Study 1.

Computation of intraindividual mean. Following data preparation procedures, mean cognitive speed was computed in the traditional manner as the mean RT of each individual's latencies across all trials for each task. Next, *intraindividual mean* or IM composites were calculated according to the same 4 factors described in Study 1 (i.e., Motor, Basic, Complex, and Interference).

Computation of intraindividual variability. The calculation of the ISD scores was identical to that described in Study 1. Although prior investigations into dementia and cognitive classification have also purified by cognitive status (e.g., Hultsch et al., 2000), the wide range of factors involved in Study 2 (e.g., MMSE change, cognitive change) would require multiple purifications according to those groupings. Further, many of the groupings are based on changes over time, precluding the ability to use them for yearly ISD calculations. Finally, although other factors such as age group are included in the regression equation, the largest systematic effects are found in relation to trial and interactions with trial. In other words, purification by group causes minimal changes to

the resulting residuals[12]. The creation of the ISD composites was identical to that described in Study 1, and formed the 4 composites corresponding to those produced for mean performance (i.e., Motor, Basic, Complex, and Interference).

Statistical Analyses.

Multivariate analyses of variance were used to establish the 5-year longitudinal link between cognitive performance and inconsistency. Hierarchical regressions were used to compare this relationship with that between cognitive performance and the intraindividual mean.

Logistic regression was used to evaluate questions 1, 2, and 3 for hypotheses b) through d) because of its ability to predict discrete outcomes from a set of continuous variables with no assumptions about the distributions of the predictor variables. However, other assumptions had to be met (Tabachnick & Fidell, 2001). First, there could not be too few cases relative to the number of predictor variables. Given the large sample size, the present study satisfied this requirement. Next, to determine if there was adequate power for the goodness-of-fit tests, we evaluated the expected frequencies of the outcome variables by age group at Wave 1. According to Tabachnick and Fidell (2001), all expected frequencies should be greater than one, and no more than 20% should have an expected frequency of less than five. The expected frequencies for the cognitive change groups, the MMSE change groups, and the CIND change groups were all greater than 5, thus meeting this criterion. Third, logistic regression assumes that a linear relationship exists between the continuous predictors and the logit transformation of the

[12] An unpublished study by Hunter and Bielak (2005) compared various purification techniques (i.e., by trial; by group; trial and group), and found nearly identical results across the various techniques. They concluded that the exact purification factors have very little impact on the final results, but should always include trial to remove practice effects.

outcome variables. This was tested with the Box-Tidwell approach where interactions were formed by multiplying the 4 ISD composites and the 4 IM composites by their natural logs. These interactions and the initial predictor variables were then included in logistic regression analyses with the various group outcomes. With 16 predictors, the criterion for significance was reduced to $\alpha = .003$. There was a violation of the linearity of the logit assumption for Motor IM when evaluating word recall change ($p = .000$). A model was run which replaced Motor IM with its natural logarithm, but the assumption was still violated. However, later analyses of both Motor IM and the log of Motor IM predicting word recall change were not significant, rendering this violation harmless.

Logistic regression is also sensitive to strong multicollinearity among the predictor variables. The ISD composites were expected to significantly correlate with one another, but these correlations were moderate, ranging from .28 to .64. The correlations for the IM composites were similar, ranging from .25 to .57, indicating the assumption of an absence of high multicollinearity was met. However, because ISD and IM were derived from the same data, the correlations between the ISD and IM composites were higher. The correlations for the motor and basic factors were moderate ($r = .27$, $r = .66$), but were quite high for the complex ($r = .90$) and interference factors ($r = .82$). Given that it is impossible to solve the high multicollinearity between the IM and ISD composites, there were no analyses that directly involved both the of the corresponding IM and ISD composites (e.g., complex RT, complex ISD). Rather, the overall models of all ISD composites were compared with the overall models of the IM composites, and each composite was also evaluated individually.

Finally, both assumptions regarding the removal of outliers and the independence of responses among cases were met. Upper and lower outliers were removed from the

continuous predictors during the calculation of the IMs and ISDs, and only one case existed per participant, ensuring independence of responding.

Survival analysis was used to evaluate the attrition outcome. Survival analysis is the longitudinal version of logistic regression, and has the advantage of not only providing the knowledge of *if* an outcome is likely to occur (akin to logistic regression), but also insight into *when* the outcome might occur. Certain assumptions had to be met in order for the survival analyses to be valid, including that each individual was only represented once in the data set, the event represented a change from one of two mutually exclusive states to another (e.g., being in the study to dropping out), that all individuals had not experienced the event when they entered the study, continuous data was used, and that censoring (i.e., according to the available data, the event did not occur for an individual) was independent of the probability of the event occurrence (i.e., censoring was random or the study ended before the event occurred). The present study met all of these assumptions. Further, because the predictor variables were continuous (i.e., ISD and IM composites), and we wanted to use more than one predictor, survival regression analysis, or Cox regression was used. Cox regression requires two assumptions to be met in addition to those required for survival analysis: the sample had to be large, and the risk of event occurrence had to be constant across all time points (proportional-hazards assumption). Therefore, in the present case, the risk of drop-out had to be similar across time. The survival plot from the Kaplan-Meier method was similar to that from the Cox regression analysis, indicating the assumption had been met (Wright, 2000). However, for multivariable Cox regression, the proportional-hazards assumption had to also be met for each predictor variable, indicating that the effect of the predictor variable did not change over time. To test this assumption, a model including the interaction between

each ISD predictor and time, age group and time, and each main effect was fit to the data. An identical model substituting the interactions between each IM predictor and time, and the IM main effects was also run. Because the interactions with motor ISD, basic ISD, motor IM, interference IM, and age group were significant (*ps* < .01), the proportional-hazards assumption was violated for these measures, indicating a single hazard ratio could not be used to describe the predictor's relation to the attrition of participants. However, Wright (2000) noted that Cox regression can also be used to describe the nonproportionality of predictors by interpreting the main effect in light of the significant interaction. For example, a one unit increase in ISD results in a certain greater rate (hazard) of dropping out of the study, but the interaction with time also indicates that the hazard decreases (or increases) over time. Therefore, Cox regression was still used to address the questions in predicting attrition.

Results

The results are presented in six main parts. The first 3 questions of interest are presented in the order of each of the five cognitive outcomes [i.e., a) level of cognitive performance at Wave 6, through e) attrition]. However, to provide a complete picture for each outcome, the questions pertaining to each outcome are addressed sequentially [e.g., 1.a), 2. a), 3. a)], before moving on to the next cognitive outcome [e.g., 1. b), 2. b), 3. b)]. First, we investigate the ability of intraindividual variability to predict later cognitive ability, and meaningful cognitive change groups, before comparing the ability of the intraindividual mean to predict the same outcomes. Next, possible differences in prediction strength due to task complexity are examined. Finally, the last section of results compares the best ISD composite with standard neuropsychological tasks in their ability to predict rate of attrition.

a) Level of Cognitive Performance at Wave 6

Multivariate analysis of variance (MANOVA) was used to evaluate whether the Wave 6 score on each of the cognitive tasks (i.e., digit symbol, letter series, word recall, similarities, vocabulary) could be predicted by the various intraindividual variability composites and age group at Wave 1. A MANOVA was chosen to control for type I error. First, the overall set of independent variables from Wave 1 (i.e., ISDs and age group) significantly predicted the overall set of dependent variables (i.e., cognitive scores) at Wave 6, Pillai's trace = 1.40, $F(30, 1030) = 13.30$, $p < .001$. Next, the relationship between each predictor and the full spectrum of cognitive tasks was evaluated. Motor ISD, $F(5, 202) = 2.69$, $p < .05$, partial $\eta^2 = .06$, Basic ISD, $F(5, 202) = 2.95$, $p < .05$, partial $\eta^2 = .07$, and Complex ISD, $F(5, 202) = 4.46$, $p < .01$, partial $\eta^2 = .10$, all made significant unique contributions to predicting the entire set of cognitive measures. Age group at Wave 1 was marginally significant, $F(5, 202) = 2.07$, $p = .07$, partial $\eta^2 = .05$, but Interference ISD was not significant over and above the other predictors. Finally, we evaluated whether the set of ISD composites and age group could predict each of the cognitive outcomes 5 years later. Table 7 illustrates that the score on each cognitive test 5 years later could be significantly predicted by the Wave 1 variability composites and age group. The group of predictors accounted for the largest proportions of variance for the digit symbol, letter series, and word recall tasks. Relatively smaller, but significant amounts of variance were predicted for the vocabulary and similarities scores. Therefore, inconsistency predicted cognitive ability 5 years later regardless of the type of cognitive task. Nearly all of the relationships between the ISD composites and

Table 7

Univariate Test Results of Year 1 ISD Composites and Age Group Predicting Year 6 Cognitive Test Score

	Cognitive Test				
	Digit Symbol	Letter Series	Word Recall	Similarities	Vocabulary
Predictor	F, partial η^2	F, partial η^2	F, partial η^2	F, partial η^2	F, partial η^2
Motor	ns	ns	ns	ns	9.76**, .05
Basic	11.24**, .05	ns	6.65*, .03	ns	ns
Complex	7.60**, .04	18.43***, .08	4.13*, .02	6.46**, .03	ns
Interference	ns	5.21*, .03	ns	ns	ns
Age group	7.30**, .03	ns	ns	ns	ns
Overall	23.06***, .36	18.79***, .31	12.47***, .23	5.81***, .12	4.76***, .10

Notes. ISD composite: $F(1, 206)$; Overall: $F(5, 206)$, overall η^2. ISD = Intraindividual standard deviation.

*$p < .05$; **$p < .01$; ***$p < .001$.

cognitive score were in the expected negative direction (i.e., higher ISD, lower cognitive score), and the two values that were not negative were not significantly different from zero (i.e., 0.02).

In order to evaluate which Wave 1 IM measure was a better predictor of cognitive score 5 years later, the categorical age group variable was removed from the analyses allowing hierarchical regressions to be performed. To provide a complete understanding of the relationships between IM and ISD, the regressions were performed twice for each cognitive test: one with all ISD composites in the equation first, and the IM measures entered second; and another analysis with the IM composites entered first, and the ISD variables entered second. The respective change statistics can be found in Table 8. Overall, the unique contribution of the IM and ISD composites varied by cognitive task. The IM measures did not significantly add to the prediction of the word recall, similarities, and vocabulary scores after entry of the ISD measures. In contrast, the intraindividual variability composites did provide unique information after the entry of the intraindividual mean composites for vocabulary, and was nearly significant for word recall and similarities. On the other hand, the IM measures appeared to be a better predictor of later digit symbol score than the ISD measures, and for letter series, both the IM and ISD composites provided significant information over and above one another.

Finally, the ANOVA results presented in Table 7 were used to evaluate whether the complexity of the RT tasks each ISD variable was calculated from played a role in the ability to predict later cognitive test score. Table 9 shows the same analysis for the IM measures. In general, the greater the complexity involved in the RT tasks, the greater that composites' likelihood of uniquely contributing to predicting the cognitive score at Wave 6. Complex ISD was the best predictor, providing unique variance for 4 out of the 5

Table 8

Hierarchical Regressions of ISD and IM Composites Predicting Year 6 Cognitive Test Score

	Cognitive Test				
	Digit Symbol	Letter Series	Word Recall	Similarities	Vocabulary
Comparison 1					
ISD composites	$F(4, 207) = 26.21^{***}$	$F(4, 207) = 23.20^{***}$	$F(4, 207) = 15.47^{***}$	$F(4, 207) = 7.13^{***}$	$F(4, 207) = 5.82^{***}$
	$R^2 = .34$	$R^2 = .31$	$R^2 = .23$	$R^2_{change} = .12$	$R^2 = .10$
IM composites	$F(4, 203) = 4.47^{**}$	$F(4, 203) = 6.05^{***}$	ns	ns	ns
	$R^2_{change} = .05$	$R^2_{change} = .07$			
Comparison 2					
IM composites	$F(4, 207) = 30.26^{***}$	$F(4, 207) = 25.87^{***}$	$F(4, 207) = 14.14^{***}$	$F(4, 207) = 6.09^{***}$	$F(4, 207) = 3.77^{**}$
	$R^2 = .37$	$R^2 = .33$	$R^2 = .22$	$R^2 = .11$	$R^2 = .07$
ISD composites	ns	$F(4, 203) = 4.13^{**}$	*F(4, 203) = 2.30, p = .06,* $R^2_{change} = .03$	*F(4, 203) = 2.31, p = .06,* $R^2_{change} = .04$	$F(4, 203) = 2.55^{*}$
		$R^2_{change} = .05$			$R^2_{change} = .05$

Notes. ISD = Intraindividual standard deviation; IM = Intraindividual mean. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Table 9

Univariate Test Results of Year 1 IM Composites and Age Group Predicting Year 6 Cognitive Test Score

	Cognitive Test				
	Digit Symbol	Letter Series	Word Recall	Similarities	Vocabulary
Predictor	*F*, partial η^2	*F*, partial η^2	*F*, partial η^2	*F*, partial η^2	*F*, partial η^2
Motor	ns	ns	ns	ns	ns
Basic	10.83**, .05	ns	ns	ns	ns
Complex	19.82**, .09	34.61***, .14	17.84*, .08	8.31**, .04	9.20**, .04
Interference	ns	ns	ns	ns	ns
Age Group	12.51***, .06	4.11*, .02	ns	ns	ns
Overall	28.06***, .41	21.80***, .35	11.98***, .23	4.85***, .11	3.00*, .07

Notes. IM composite: $F(1, 206)$; Overall: $F(5, 206)$, overall η^2. ISD = Intraindividual standard deviation.

*$p < .05$; **$p < .01$; ***$p < .001$.

cognitive tasks. The IM measures concurred with this conclusion and provided even clearer superiority of the complex composite in predicting later cognitive test score.

Cognitive Outcomes b) Through d)

A series of multinomial logistic regression analyses were used to address questions 1, 2, and 3 for the group outcomes b) change in cognitive performance, c) MMSE change, and d) cognitive status change. Question 1, whether inconsistency predicted the relevant outcome group, was evaluated using a model including all ISD composites and age group at Wave 1. Question 2, addressing whether ISD could predict the group outcomes better than IM involved a model including all IM composites and age group at Wave 1, and compared the IM model to the ISD model. Question 3, which asked whether the complexity of the RT tasks played a role in prediction, was evaluated using individual multinomial logistic regression for each predictor. The following tests of significance were adjusted for type I error by dividing $\alpha = .05$ by the number of predictors in the model, resulting in critical values of $\alpha = .01$ for the 5 predictor ISD and IM models. Summaries of the odds ratios predicted by each of the ISD and IM composites across the 5 cognitive tasks are provided in Tables 10 and 11, respectively, and a summary for the CIND change group can be found in Table 12.

When interpreting the odds ratios produced by the logistic regression analyses however, it is important to note that the scales of the ISD and IM measures differed. A unit increase in ISD was one tenth of a standard deviation in variability, because the residuals were converted to *T*-scores before the individual standard deviation was calculated. On the other hand, a one unit increase in IM was one millisecond of change in reaction time. Consequently, significantly smaller odds ratios were expected for the IM measures compared to those associated with the ISD measures. To enhance interpretation

Table 10

Odds Ratios of Year 1 ISD Composites Individually Predicting Cognitive Test Change Groups

	Cognitive Test				
	Digit Symbol	Letter Series	Word Recall	Similarities	Vocabulary
ISD Composite	Exp(β)	Exp(β)	Exp(β)	Exp(β)	Exp(β)
Motor	D vs. I = .801* *S vs. I = .865, p = .09*	ns	D vs. I = .774*	ns	ns
Basic	ns	ns	D vs. S = .587*** D vs. I = .510***	ns	*D vs. I = .772, p = .07*
Complex	ns	ns	D vs. S = .720*** D vs. I = .641***	ns	ns
Interference	*D vs. S = .898, p = .09* *D vs. I = .888, p = .10*	ns	D vs. S = .859* D vs. I = .815*	ns	ns

Notes. ISD = Intraindividual standard deviation; D = Decline; S = Stable; I = Improve. The reference group is listed first. 1 unit = 0.1 standard deviation. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 11

Odds Ratios of Year 1 IM Composites Individually Predicting Cognitive Test Change Groups

	Cognitive Test				
	Digit Symbol	Letter Series	Word Recall	Similarities	Vocabulary
IM Composite	Exp(β)	Exp(β)	Exp(β)	Exp(β)	Exp(β)
Motor	ns	ns	ns	ns	*D vs. I = .9879, p = .09*
Basic	ns	ns	D vs. S = .9919** D vs. I = .9883***	ns	ns
Complex	ns	ns	D vs. S = .9980*** D vs. I = .9971***	ns	ns
Interference	ns	ns	D vs. S = .9962* D vs. I = .9950*	*D vs. I = .9962, p = .10*	ns

Notes. IM = Intraindividual mean; D = Decline; S = Stable; I = Improve. The reference group is listed first. 1 unit = 1 millisecond.

*$p < .05$; **$p < .01$; ***$p < .001$.

Table 12

Odds Ratios of Year 1 ISD and IM Composites Individually Predicting CIND Change Groups

	RT Measure	
	ISD	IM
Composite	Exp(β)	Exp(β)
Motor	*St. intact vs. St. CIND = 1.224, p = .05*	*St. intact vs. St. decline = 1.0128, p = .08*
Basic	St. intact vs. Fluctuating = 1.290* St. intact vs. St. decline = 1.609** St. intact vs. St. CIND = 1.398*	*St. intact vs. Fluctuating = 1.0034, p = .10* *St. intact vs. St. decline = 1.0060, p = .08* St. intact vs. St. CIND = 1.0062*
Complex	St. intact vs. Fluctuating = 1.197** St. intact vs. St. decline = 1.324** St. intact vs. St. CIND = 1.378**	St. intact vs. Fluctuating = 1.0012** St. intact vs. St. decline = 1.0016* St. intact vs. St. CIND = 1.0023*** *Fluctuating vs. St. CIND = 1.0011, p = .08*
Interference	St. intact vs. Fluctuating = 1.135* St. intact vs. St. decline = 1.229**	*St. intact vs. St. CIND = 1.0040, p = .06*

Notes. ISD = Intraindividual standard deviation (1 unit = 0.1 SD); IM = Intraindividual mean (1 unit = 1 msec); St. = Stable. The reference group is listed first. *$p < .05$; **$p < .01$; ***$p < .001$.

of the odds ratios, the ISD odds ratios will be multiplied by 10, resulting in the odds ratio for the more meaningful 1.0 standard deviation change in ISD. Similarly, the more meaningful 100 millisecond change will be used when interpreting the IM result by multiplying the odds ratios by 100. The multiplication procedures will only be employed when making inferential statements about the odds ratios, following the description of the original statistics in the text and tables.

b) Change in Cognitive Performance

Digit Symbol

The model with age group at Wave 1 and the motor, basic, complex, and interference ISDs was not significant in distinguishing among the digit symbol change groups (stable, improve, decline). Further, none of the variables were noted as having a significant effect on the model if removed, and none of the Wald comparisons were significant.

A comparison model with all IM measures and age group at Wave 1 revealed similar results. The overall model, the tests of removing each variable, and the Wald tests were all not significant. However, the χ^2 value, (10, N = 212) = 8.36, and the related Nagelkerke's R^2 = .05, were significantly smaller than the model with all 4 ISD variables and age group, χ^2 (10, N = 212) = 12.92, Nagelkerke's R^2 = .07, indicating a trend towards the intraindividual variability measures being better predictors of digit symbol change group.

To evaluate whether complexity of the RT task played a role in each measure's ability to predict change group status, models with each individual ISD or IM measure were run. Motor ISD was the only intraindividual variability measure to even approach significance in differentiating among the digit symbol change groups, χ^2 (2, N = 212) =

5.02, $p = .08$, Nagelkerke's $R^2 = .03$. Motor ISD significantly distinguished between the improve group and the decline group, $\beta = -.222$, SE = .11, Wald (1) = 4.29, $p < .05$, Exp(β) = .801, where a one SD increase in motor ISD reduced an individual's likelihood of being in the improve group than the decline group by 199%. Further, motor ISD was almost able to significantly distinguish the improve group from the stable group, β = -.145, SE = .08, Wald (1) = 2.96, $p = .09$, Exp(β) = .865, in that a one standard deviation increase in motor ISD decreased an individual's chances of being in the improve group compared to the stable group by 135%. Although the overall model with interference ISD was not significant, χ^2 (2, $N = 212$) = 3.20, Nagelkerke's $R^2 = .02$, the Wald comparisons approached significance: Decline versus stable group, β = -.108, SE = .06, Wald (1) = 2.86, $p = .09$, Exp(β) = .898, and Decline versus improve group, β = -.119, SE = .07, Wald (1) = 2.72, $p = .10$, Exp(β) = .888. A one SD increase in interference ISD almost significantly resulted in a 102% decreased likelihood of being in the stable digit symbol change group, and a 112% decreased likelihood of being in the improve digit symbol change group, compared to being in the decline group. In contrast, none of the mean measures were individually significant, nor were their related Wald comparison tests (see Tables 10 and 11). Overall, greater complexity of the RT tasks did not enhance prediction of the digit symbol change group. Rather, the simplest and intermediately complex of the four composites were the best individual predictors of later group status, but only among the ISDs. The IM measures were all poor predictors, regardless of task difficulty.

As an aside, age group at Wave 1 was able to significantly differentiate among the digit symbol groups, χ^2 (2, $N = 212$) = 6.40, $p < .05$, Nagelkerke's $R^2 = .04$, specifically the improve group from the stable, β = .77, SE = .35, Wald (1) = 4.84, $p < .05$, Exp(β) =

2.16, and decline groups, β = .99, SE = .47, Wald (1) = 4.38, p <.05, Exp(β) = 2.68. A member of the young-old age group was 2.16 times, or 116% more likely to be in the improve group than the stable group, and 2.68 times, or 168% more likely to be in the improve group than the decline group.

Letter Series

As a group, the 4 ISD composites and age group at Wave 1 did not make a significant difference in predicting letter series change group over the 5 years. However, the removal of Interference ISD was almost statistically significant (with 5 predictors, α = .01), χ^2 (2, N = 212) = 6.17, p = .05, as was its related Wald test comparing the decline and stable groups, β = .174, SE = .09, Wald (1) = 3.67, p = .06, Exp(β) = 1.190, and the stable and improve groups, β = -.175, SE = .10, Wald (1) = 3.07, p = .08, Exp(β) = .840. For each SD increase in interference ISD score at Wave 1, there was an unexpected 190% increased likelihood of being in the letter series stable group compared to the decline group, and a 160% decreased likelihood of being in the improve compared to the stable change group.

A model with all IM measures and age group revealed similar results. The overall model was not significant, but the test of removing interference IM was nearly significant, χ^2 (2, N = 212) = 6.40, p = .04, and less so, basic IM, χ^2 (2, N = 212) = 4.67, p = .10. There were corresponding nearly significant Wald tests for each variable. For interference IM, the comparison between the stable and decline groups, β = -.0049, SE = .002, Wald (1) = 3.95, p = .05, Exp(β) = .9951, and between the stable and improve groups, β = -.0045, SE = .003, Wald (1) = 2.92, p = .09, Exp(β) = .9955, approached significance. Similarly, for basic IM the comparison between the decline and stable groups neared significance, β = -.0056, SE = .003, Wald (1) = 3.08, p = .08, Exp(β) =

.9944. For letter series, each 100 millisecond increase in responding to interference RT measures unexpectedly increased a participant's likelihood of being in the stable group versus the decline group by 49%, but decreased their likelihood of being in the improve group compared to the stable group by 45%. Longer responding on the Basic RT tasks was also detrimental, and decreased the likelihood of being in the stable letter series group compared to the decline group by 56%. Although the χ^2 value for the mean predictors and age group, (10, N = 212) = 11.89, was smaller than that for the ISD predictors and age group, χ^2 (10, N = 212) = 12.59, this difference was negligible, as also indicated by the nearly identical Nagelkerke's R^2 values (.067 and .071, respectively). Overall, the lack of full model significance, and marginally significant individual predictor results suggested that the ISD and IM measures were equally poor at distinguishing among the letter series change groups.

In terms of enhanced prediction based on the complexity of the RT tasks, individual models with each ISD or IM measure were analyzed. However, none of the individual ISD or IM models were significant, or included significant Wald comparisons. Therefore, the complexity of the RT task the variability and mean measures were based on did not play a role in predicting change group status on the letter series task.

Word Recall

The full model test of age group at Wave 1, motor ISD, basic ISD, complex ISD, and interference ISD was significant, χ^2 (10, N = 212) = 35.48, $p < .001$, Nagelkerke's R^2 = .18, demonstrating that the group of predictors significantly differentiated between the improve, decline, and stable word recall change groups. Basic ISD was the only predictor whose removal was marginally significant, χ^2 (2, N = 212) = 7.18, p = .03. Further, the related Wald comparisons for basic ISD were almost significant in distinguishing the

decline group from the stable, β = -.382, SE = .15, Wald (1) = 6.20, *p* =.01, Exp(β) = .682, and improve groups, β = -.427, SE = .19, Wald (1) = 5.01, *p* = .03, Exp(β) = .653. Both comparisons were in the expected direction, in that a one standard deviation increase in basic ISD during Wave 1 resulted in a 318% less likelihood of being in the stable group than the decline group, and a 347% less likelihood of being in the improve group than the decline group. Interestingly, although complex ISD was not noted as causing a significant distortion in prediction if removed from the model, change in the amount of complex ISD was almost significant in promoting distinction between the decline and improve groups, β = -.267, SE = .14, Wald (1) = 3.44, *p* =.06, Exp(β) = .766. In this case, greater inconsistency in completing the RT tasks for the complex composite by one standard deviation resulted in a 234% reduced likelihood of being in the improve change group for word recall compared to the decline group.

On the other hand, the full model of all IM predictors and age group at Wave 1 was also significant in predicting the word recall change groups, χ^2 (10, N = 212) = 32.12, $p < .001$, Nagelkerke's R^2 = .17. Complex IM was the only measure whose removal would almost significantly impact the model, χ^2 (2, N = 212) = 7.06, p = .03. Its related Wald comparisons revealed complex IM was marginally significant in differentiating among the decline and stable, β = .0013, SE = .001, Wald (1) = 4.50, *p* =.03, Exp(β) = 1.0013, and improve groups, β = -.0020, SE = .001, Wald (1) = 6.17, *p* =.01, Exp(β) = .9980. Therefore, an one hundred millisecond increase for complex IM resulted in a 13% increased likelihood of being in the decline word recall group compared to the stable group, and a 20% decreased likelihood of being in the improve group compared to the decline group. Comparisons for basic IM were similarly approaching significance for distinguishing between the decline and stable groups, β = .0057, SE = .003, Wald (1) =

3.02, p =.08, Exp(β) = 1.0057, and the decline and improve groups, β = -.0081, SE = .004, Wald (1) = 3.83, p =.05, Exp(β) = .9919, but the associated risks were greater. For each one hundred more milliseconds in basic IM, the likelihood of being in the decline group rather than the stable group increased by 57%, and the likelihood of being in the improve group rather than the decline group decreased by 81%. Age group was also nearly significant: decline versus stable, β = .90, SE = .47, Wald (1) = 3.62, p =.06, Exp(β) = 2.45, indicating that those in the young-old age group were 145% more likely to be in the stable than decline change group. Although both models with the IM and ISD measures were significant at predicting word recall change group, the model involving the ISDs was marginally stronger, χ^2 = 35.48, Nagelkerke's R^2 = .18, than that with the IMs, χ^2 = 32.12, Nagelkerke's R^2 = .17. Overall however, it appears that the ISD and IM measures are comparable in differentiating among the word recall 5-year change groups.

The individual contributions of the ISD and IM composites to predicting word recall change group were substantial. All but two of the individual models resulted in a significant overall model, complicating the investigation into the possible impact of cognitive complexity. However, there were clear differences among the variables in terms of the strength of the predictive relationship. For the ISDs, basic ISD, χ^2 (2, N = 212) = 26.22, p < .001, Nagelkerke's R^2 = .14, and complex ISD, χ^2 (2, N = 212) = 25.77, p < .001, Nagelkerke's R^2 = .14, were the strongest predictors, distantly followed by interference ISD, χ^2 (2, N = 212) = 8.30, p < .05, Nagelkerke's R^2 = .05, and the marginally significant motor ISD, χ^2 (2, N = 212) = 5.48, p =.07, Nagelkerke's R^2 = .03. For basic ISD, there were significant distinctions between the decline group and the stable, β = -.532, SE = .12, Wald (1) = 18.41, p <.001, Exp(β) = .587, and improve groups, β = -.674, SE = .15, Wald (1) = 19.17, p <.001, Exp(β) = .510. For each standard

deviation increase in basic ISD at Wave 1, an individual was 413% less likely to be in the word recall stable group compared to the decline group, and 490% less likely to be in the word recall improve group than the decline group. The complex ISD measure yielded similar but not as strong odds comparisons. For the odds ratio provided by Complex ISD and the other ISD measures, see Table 10.

For the IMs, only motor IM did not independently predict change group. Compared to the ISDs however, it was complex IM rather than basic IM that provided the strongest prediction [χ^2 (2, N = 212) = 24.61, p <.001, Nagelkerke's R^2 = .13, and χ^2 (2, N = 212) = 16.30, p <.001, Nagelkerke's R^2 = .09, respectively], and interference IM was distant third, χ^2 (2, N = 212) = 6.39, p <.05, Nagelkerke's R^2 = .04. However, the odds ratios comparing the decline, stable, and improve groups were greatest for the basic IM measure. The statistics for differentiating between the decline and stable groups based on basic IM score were, β = -.0082, SE = .003, Wald (1) = 10.24, p <.01, Exp(β) = .9919, indicating one hundred more milliseconds in basic IM decreased the likelihood of being in the stable group compared to the decline group by 81%. Even stronger differentiation was found between the decline and improve groups, β = -.0118, SE = .003, Wald (1) = 13.79, p <.001, Exp(β) = .9883, as the likelihood of being in the improve group was reduced by an impressive 117%. Odds ratios for the other IMs can be found in Table 11. Overall, it appeared that the complexity of the RT task did play a role in the Wave 1 intraindividual variability or mean measures being able to predict word recall change group. Specifically, ISD and IM values based on moderately complex RT tasks provided the most insight.

Finally, age group at Wave 1 also reliably distinguished the groups, χ^2 (2, N = 212) = 12.70, p < .01, Nagelkerke's R^2 = .07, specifically the decline group from the

stable, β = 1.33, SE = .40, Wald (1) = 10.93, p <.001, Exp(β) = 3.79, and improve groups, β = 1.44, SE = .49, Wald (1) = 8.70, p <.01, Exp(β) = 4.23. If a participant was a member of the young-old age group at Wave 1, they were 279% more likely to be in the stable group than the decline group, and 323% more likely to be in the improve group.

Similarities

As a group, the 4 ISD composites and age group at Wave 1 did not make a significant difference in predicting similarities 5-year change group. Further, none of the variables were noted as causing a significant distortion in prediction if removed from the model, and none of the Wald comparisons were significant.

A comparison model with all IM measures and age group also failed to distinguish among the similarities change groups. Although the test of removing interference IM was far from being significant given the 5 predictors in the model (α = .01), χ^2 (2, N = 212) = 4.76, p = .09, there were corresponding nearly significant Wald tests. Interference IM appeared nearly able to significantly differentiate the decline group from the stable group, β = -.0042, SE = .002, Wald (1) = 3.90, p =.05, Exp(β) = .9958, and the decline group from the improve group, β = -.0054, SE = .003, Wald (1) = 3.50, p =.06, Exp(β) = .9947. Compared to the decline group on the similarities task, individuals had a 42% reduced likelihood of being in the stable change group, and a 53% reduced likelihood of being in the improve change group, for each 100 milliseconds slower responding in interference IM. Although the ISD composites model and the IM composites model were both nonsignificant, the models were nearly identical: χ^2 (10, N = 212) = 7.68, Nagelkerke's R^2 = .04; χ^2 (10, N = 212) = 8.49 Nagelkerke's R^2 = .05, respectively. As a result, it appeared that neither RT-derived measure was particularly adept at predicting similarities change group status.

In terms of enhanced prediction based on the complexity of the RT tasks, individual models with each ISD or IM measure were analyzed. Interference IM was the only model that, while nonsignificant overall, had a nearly significant Wald comparison in distinguishing between the decline and improve groups, β = -.0038, SE = .002, Wald (1) = 2.79, *p* =.10, Exp(β) = .9962. The interpretation of the related odds ratio meant each 100 millisecond slower responding on interference IM was related to a 38% reduced likelihood of being in the improve change group versus the decline group. Overall, the complexity of the RT task the variability and mean measures were based on did not play a role in predicting change group status on the similarities task.

Vocabulary

The full model test of age group at Wave 1, motor ISD, basic ISD, complex ISD, and interference ISD was not significant, but interference ISD was the only measure whose removal would almost significantly impact the model's prediction, χ^2 (2, N = 212) = 5.45, p = .07. Further, interference ISD score was found to nearly significantly differentiate between the stable and improve vocabulary change groups, β = .194, SE = .08, Wald (1) = 5.34, *p* =.02, Exp(β) = 1.214, in that a one SD increase in interference inconsistency would increase an individual's likelihood of improving in vocabulary over the 5 years by 214%, compared to remaining stable.

The model of IM composites and age group similarly failed to significantly predict the change groups, but the removal of interference IM would nearly significantly affect the model, χ^2 (2, N = 212) = 6.57, p = .04. Similar to its ISD measure, interference IM was marginally significantly able to distinguish among the stable and improve groups, β = .0054, SE = .002, Wald (1) = 6.25, *p* =.01, Exp(β) = 1.0054, but was also marginally significantly able to differentiate among the decline and improve groups, β = .0055, SE =

.003, Wald (1) = 3.67, p =.06, Exp(β) = 1.0055. Therefore, for each 100 milliseconds slower a participant responded to the interference RT tasks, they were 54% more likely to be in the improve change group compared to the stable group, and similarly 55% more likely to be in the improve group compared to the decline group. Comparing the overall model for ISD, χ^2 (10, N = 212) = 13.19, Nagelkerke's R^2 = .07, with that for IM, χ^2 (10, N = 212) = 11.61, Nagelkerke's R^2 = .07, both of which were nonsignificant, revealed no apparent superiority of either RT measure in predicting vocabulary change group.

Independently, none of the ISD or IM models were significant. Motor IM had a nearly significant Wald comparison between the decline and improve groups, β = -.0122, SE = .007, Wald (1) = 2.86, p =.09, Exp(β) = .9879, indicating participants were 121% more likely to be in the decline group than the stable group for each one hundred millisecond increase in Motor IM. Basic ISD was almost significant in distinguishing between the decline and improve groups, β = -.259, SE = .14, Wald (1) = 3.31, p =.07, Exp(β) = .772, indicating that each standard deviation increase in basic ISD at Wave 1 reduced the likelihood of being in the improve group compared to the decline vocabulary change group by 228%. Overall, the few and only nearly significant individual results across both the ISD and IM measures suggested that the complexity of the RT task made little difference in being able to predict 5-year change group on the vocabulary task. However, there was a trend towards moderately difficult tasks playing a greater role among the inconsistency measures.

c) MMSE Change

As a set, the various Wave 1 ISD composites and age group failed to significantly differentiate among the 3 MMSE five-year change groups (improve, stable, decline). None of the variables were noted as causing a significant distortion in prediction if

removed from the model, but the Wald comparison between the stable and improve groups was nearly significant for interference ISD, β = -.156, SE = .08, Wald (1) = 3.57, *p* =.06, Exp(β) = .856. A one standard deviation increase in interference inconsistency would reduce an individual's likelihood of being in the improve change group rather than the stable change group by 144%. The Wald comparison for the decline and stable groups was nearly significant for age group as well (*p* = .05). However, given that a later model with only age group was significant and provided a stronger odds ratio, the Wald statistics and odds ratios from that analysis will be described in the individual comparisons.

The model with all IM composites and age group at Wave 1 also failed to significantly distinguish among the MMSE change groups. Age group was the only predictor noted as nearly causing a significant impact if removed from the model, χ^2 (2, *N* = 212) = 5.79, *p* = .06, and age group's ability to distinguish among the decline and stable groups was almost significant (*p* = .02). The Wald statistics of a model with only age group will be described shortly. The model comprising of mainly ISD composites and that comprising of mainly IM composites were both not significant, and their corresponding χ^2 values and effect sizes were practically identical (χ^2 = 12.98 and 13.96, respectively, Nagelkerke's R^2 = .08 for both), indicating neither RT measure was better at predicting the MMSE change groups than the other.

Although none of the individual ISD or IM models were significant overall, the complex ISD Wald comparison distinguishing the decline and stable groups was significant, β = -.184, SE = .09, Wald (1) = 3.86, *p* <.05, Exp(β) = .832. Therefore, a one standard deviation increase in ISD on the complex RT tasks at Wave 1 reduced the likelihood of being in the stable MMSE change group 5 years later compared to being in

the decline MMSE change group by 168%. Given that complex ISD was the only significant model, it appeared that RT tasks with the highest amounts of cognitive complexity offered the best predictability of MMSE change over 5 years. However, the complexity of the task appeared irrelevant for IM in predicting MMSE change group.

As an aside, age group at Wave 1 significantly predicted MMSE change group, χ^2 (2, N = 212) = 6.62, p <.05, Nagelkerke's R^2 = .04, specifically distinguishing between the stable and decline groups, β = 1.31, SE = .54, Wald (1) = 5.96, p <.05, Exp(β) = 3.70. If a participant was in the young-old age group, they were 270% more likely to be in the MMSE stable group than the MMSE decline group.

d) Cognitive Status Change

As a group, Wave 1 motor, basic, complex, and interference ISD and age group were able to significantly predict CIND change group status, χ^2 (15, N = 212) = 40.69, p < .001, Nagelkerke's R^2 = .20. Age group and complex ISD were noted as being particularly important to the model, as removal of age group would significantly reduce the model's power, χ^2 (3, N = 212) = 16.16, p < .01, and removal of complex ISD neared significance, χ^2 (3, N = 212) = 9.57, p < .05. Given that age group is not the main focus of the present analyses, its significance and various group comparisons will be noted in the individual model comparisons. Complex ISD was able to significantly distinguish between the stable CIND and stable intact groups, β = -.399, SE = .15, Wald (1) = 7.44, p <.01, Exp(β) = .671, and almost significantly differentiate between stable CIND and stable decline, β = -.342, SE = .19, Wald (1) = 3.14, p =.08, Exp(β) = .710, and fluctuating and stable intact, β = -.206, SE = .11, Wald (1) = 3.82, p =.05, Exp(β) = .814. Therefore, for each SD increase in complex ISD, an individual was 329% less likely to be in the stable intact group, and 290% less likely to be in the stable decline group,

compared to the stable CIND group. Further, an individual was 186% less likely to be in the stable intact group compared to the fluctuating group.

In comparison, the model with all 4 IM composites and age group at Wave 1 was also significant in distinguishing among the various CIND change groups, χ^2 (15, N = 212) = 34.69, $p < .01$, Nagelkerke's $R^2 = .17$. Following the lead of the ISD model, complex IM and age group would both cause significant disruptions to the overall group prediction if removed from the model, χ^2 (3, N = 212) = 17.56, $p < .01$, and χ^2 (3, N = 212) = 11.43, $p = .01$, respectively. Complex IM distinguished between the same groups as complex ISD in the ISD model: Stable CIND and stable intact, β = -.0027, SE = .001, Wald (1) = 11.35, $p < .01$, Exp(β) = .9973; stable CIND and stable decline, β = -.0019, SE = .001, Wald (1) = 3.11, $p = .08$, Exp(β) = .9981; and fluctuating and stable intact, β = -.0018, SE = .001, Wald (1) = 9.89, $p < .01$, Exp(β) = .9982. Therefore, for each one hundred millisecond increase in complex IM, there was a 27% reduced likelihood of being in the stable intact group, and 19% reduced likelihood of being in the stable decline group, compared to the stable CIND change group. Further, an individual was 18% less likely to be in the stable intact group compared to the fluctuating group. Overall, the ISD model appeared to do a slightly superior job at predicting CIND change group, given its higher χ^2 and effect size ($\chi^2 = 40.69$, $R^2 = .20$, versus $\chi^2 = 34.69$, $R^2 = .17$ for the IM model) versus that of complex IM.

Individually, there were numerous significant and almost significant models predicting CIND change group (see Table 12). For the ISDs, the basic, χ^2 (3, N = 212) = 15.24, $p < .01$, Nagelkerke's $R^2 = .08$, complex, χ^2 (3, N = 212) = 18.62, $p < .001$, Nagelkerke's $R^2 = .10$, and interference, χ^2 (3, N = 212) = 9.17, $p < .05$, Nagelkerke's R^2 = .05, composites were significant. Although the complex ISD model had the largest

effect size, the basic ISD model had stronger odds ratios in distinguishing the groups. Specifically, basic ISD was able to significantly distinguish the stable intact group from the 3 other groups: Fluctuating, β = .255, SE = .10, Wald (1) = 6.66, p <.05, Exp(β) = 1.290; stable decline, β = .476, SE = .16, Wald (1) = 8.50, p <.01, Exp(β) = 1.609; and stable CIND, β = .335, SE = .14, Wald (1) = 5.63, p <.05, Exp(β) = 1.398. Therefore, a one point increase in basic ISD score at Wave 1 increased an individual's likelihood of being in the fluctuating group by 290%, the stable decline group by 609%, and the stable CIND group by 398%. The Wald statistics and odds ratios for the other ISD measures were in the same direction, and accounted for similar group comparisons (see Table 12). Given that the strongest comparisons between the CIND change groups was from basic ISD, it appeared that the moderately complex RT tasks provided the best insight into predicting later cognitive change.

For the IMs, the complex model was significant, χ^2 (3, N = 212) = 20.59, p <.001, Nagelkerke's R^2 = .10, and the basic model was almost significant, χ^2 (3, N = 212) = 7.37, p =.06, Nagelkerke's R^2 = .04, in predicting CIND change group. Complex IM also had more significant distinctions among the groups than the other IM composites (see Table 12). Complex IM was able to significantly distinguish the stable intact group from the fluctuating group, β = .0012, SE = 0, Wald (1) = 8.21, p <.01, Exp(β) = 1.0012, the stable decline group, β = .0016, SE = .001, Wald (1) = 5.20, p <.05, Exp(β) = 1.0016, and the stable CIND group, β = .0023, SE = .001, Wald (1) = 14.47, p <.001, Exp(β) = 1.0023. There was also a marginally significant distinction between the fluctuating and stable CIND groups, β = .0011, SE = .001, Wald (1) = 3.04, p =.08, Exp(β) = 1.0011. Therefore, compared to the stable intact group, each one hundred millisecond increase in complex IM increased an individual's likelihood of being in the fluctuating group by

12%, the stable decline group by 16%, and the stable CIND group by 23%. Further, the likelihood of being in the stable CIND group increased 11% compared to being in the fluctuating change group. Consequently, it appeared that the complexity of the RT tasks was important in providing the strongest IM prediction of later CIND change group. However, in contrast to ISD where moderately complex tasks were superior, highly complex tasks provided the best differentiation for IM.

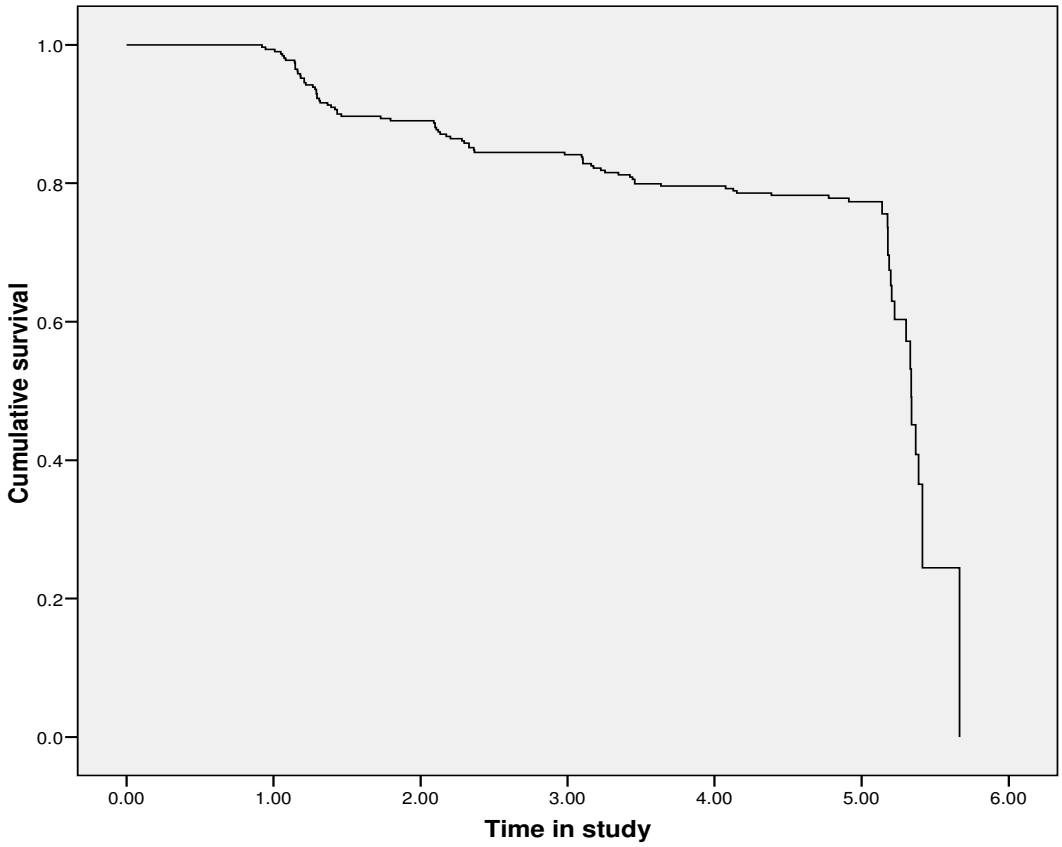
As an aside, age group at Wave 1 significantly predicted CIND change group, χ^2 (3, N = 212) = 8.01, p <.05, Nagelkerke's R^2 = .04, specifically distinguishing the stable decline group from the stable intact, β = 1.21, SE = .58, Wald (1) = 4.38, p <.05, Exp(β) = 3.36, and fluctuating groups, β = 1.66, SE = .62, Wald (1) = 7.12, p <.01, Exp(β) = 5.25. Compared to being in the stable decline group, a young-old participant was 236% more likely to be in the CIND stable group, and 425% more likely to be in the fluctuating group.

e) Attrition

First, Figure 5 shows the survival function for remaining in the study across the testing waves. It is apparent that cumulative survival dropped more between Waves 1 and 2 (i.e., between 0 and 365 days) than the later waves of testing, indicating that most participants who did not continue in the study chose to do so early on. Note that the drop in survival between years 5 and 6 is an artefact of when the study completed its last testing wave (it actually represents Wave 6 and hypothetical Wave 7). For example, participants who remained in the study had a mean time of testing of 4.93 years (i.e., close to 5 years, or 6 waves of testing), while those who dropped out had a mean time in study of 2.78 years.

Figure 5

Cumulative Survival Function of Attrition by Time in Study



Notes. The first wave of testing was considered baseline performance and noted as "0". Therefore time 0 is actually wave 1 testing, and so forth. Participants did not complete the study at equal intervals, thus resulting in some participants' time in study appearing to be past the year 6 data collection.

To determine if Wave 1 ISD was able to differentiate those who dropped out of the study from those who remained in the study, we ran a Cox regression model including all ISD composites, age group, and the significant age group, motor ISD, and basic ISD interactions with time. The overall model was significant, χ^2 (8, N = 304) = 639.78, p <.001, in predicting rate of attrition. Table 13 shows the statistics for each predictor, holding the effects of the other predictors constant. Interestingly, both complex ISD and interference ISD were the only predictors which did not contribute any unique variance to predicting attrition. Rather, for a one standard deviation increase in motor ISD at Wave 1, the risk of attrition uniquely increased by 340%, but this risk also significantly decreased slightly over time (i.e., Motor ISD X Time, β = -.116). Further, participants who showed greater inconsistency on the basic ISD task were at a substantially greater risk of dropping out of the study; with each one SD increase in inconsistency on the basic RT tasks, the rate of attriting uniquely increased by an impressive 1605%. This effect also significantly slightly decreased over time however (i.e., Basic ISD X Time, β = -.278). Finally, age group played a large role in determining later attrition. Individuals who were in the old-old age group (i.e., 75+ years) had a 15.55 times greater rate of attrition than those in the young-old age group (i.e., 65-74 years; see Figure 6), but this enhanced risk decreased over time in study (i.e., Age Group X Time, β = -.859). However, these risks varied for the population; with 95% confidence, we can infer that for the population, the risk of attrition increases between 128% and 594% for each one standard deviation increase in motor inconsistency, between 991% and 2407% for each one unit increase in basic inconsistency, and between 423% and 4525% for those in the old-old age group. Overall, increased inconsistency and older age predict later rate of attrition well.

Table 13

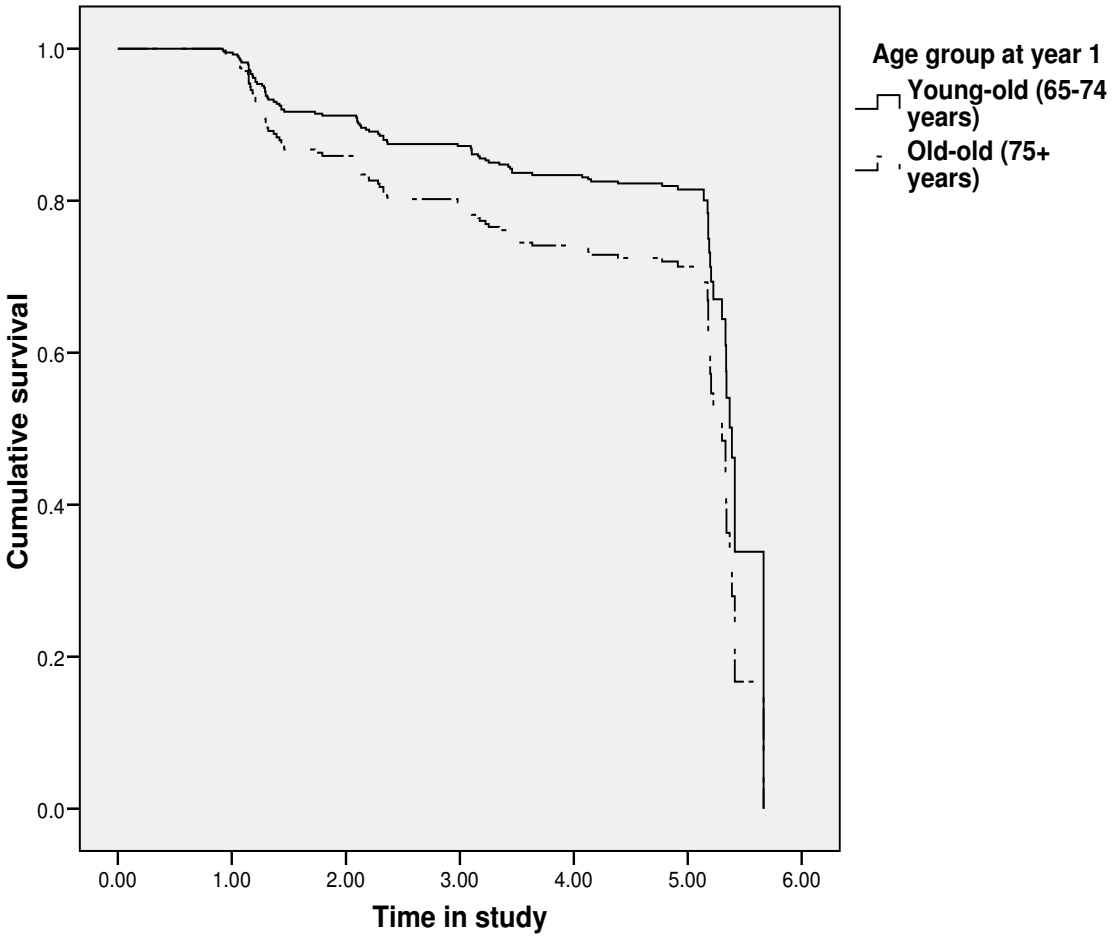
Model Statistics of Year 1 ISD Composites and Age Group Predicting Rate of Attrition, Controlling for Other Predictors

	Statistic				
Predictor	β	SE	Wald	Exp(β)	95% CI Exp(β)
Motor ISD	.293	.09	11.03**	1.340	1.128 – 1.594
Basic ISD	.957	.14	48.83***	2.605	1.991 – 3.407
Complex ISD	-.031	.07	.18	.970	.843 – 1.117
Interference ISD	-.086	.05	2.69	.918	.829 – 1.017
Age Group	2.744	.56	24.35***	15.551	5.228 – 46.254
Motor ISD X Time	-.116	.03	11.76**	.890	.833 – .951
Basic ISD X Time	-.278	.03	70.53***	.758	.710 – .808
Age Group X Time	-.859	.15	32.05***	.424	.315 – .570

Notes. ISD = Intraindividual standard deviation (1 unit = 0.1 SD); Exp(β) = Hazard ratio; CI = Confidence interval. *$p < .05$; **$p < .01$; ***$p < .001$.

Figure 6

Cumulative Survival Function of Attrition by Time in Study According to Age Group



Notes. The first wave of testing was considered baseline performance and noted as "0". Therefore time 0 is actually wave 1 testing, and so forth. Participants did not complete the study at equal intervals, thus resulting in some participants' time in study appearing to be past the year 6 data collection.

In comparison, the Cox regression model including all IM composites, age group, and the significant age group, motor IM, and interference IM interactions with time, was also significant, χ^2 (8, N = 304) = 643.26, p <.001, in predicting rate of attrition. Table 14 shows the statistics for each predictor, holding the effects of the other predictors constant. Although basic and complex IM were non-significant predictors, both Wave 1 motor and interference IM were uniquely significant in predicting rate of attrition. For each one hundred millisecond increase in motor IM at Wave 1, the risk of later dropping out of the study uniquely increased by 290%, but this risk declined over time in study (i.e., Motor IM X Time, β = -.008). The relationship between responding slower on the interference RT tasks at Wave 1 and rate of attrition were much weaker, but still uniquely significant; with each one hundred millisecond increase in IM on the interference tasks, the rate of attriting increased by 70%. This effect also abated slightly over time (i.e., Interference IM X Time, β = -.008). The age group effects were similar to those described for the ISD model. Overall, despite showing similar model significance, the IM and ISD models varied in the unique predictors of rate of attrition.

Cox regressions with individual continuous predictors were run to evaluate the hypothesis that complexity of the RT tasks may have played a role in the strength of predicting the rate of attrition. All of the ISD composites were significant predictors (*p*s < .05), but the results corresponded to those reported in the multivariable analyses: Complex and interference ISD had much smaller hazard ratios, Exp(β) = 1.116, and 1.086, compared to the motor and basic measures, Exp(β) = 2.272, and 2.436. However, the motor and basic measures were tempered by their interactions with time (β = -.30), and the hazard ratios declined over the course of the study. Overall, basic ISD offered the

Table 14

Model Statistics of Year 1 IM Composites and Age Group Predicting Rate of Attrition, Controlling for Other Predictors

	Statistic				
Predictor	β	SE	Wald	Exp(β)	95% CI Exp(β)
Motor IM	.029	.01	21.21***	1.029	1.017 – 1.042
Basic IM	.003	.002	2.67	1.003	.999 – 1.007
Complex IM	.000	.00	.09	.999	.999 – 1.001
Interference IM	.007	.002	12.34***	1.007	1.003 – 1.011
Age Group	2.41	.57	17.82***	11.17	3.64 – 34.24
Motor IM X Time	-.008	.002	18.74***	.992	.988 – .995
Interference ISD X Time	-.004	.001	28.37***	.996	.995 – .998
Age Group X Time	-.781	.16	23.45***	.458	.334 – .628

Notes. IM = Intraindividual mean (1 unit = 1 msec); Exp(β) = Hazard ratio; CI = Confidence interval. *$p < .05$; **$p < .01$; ***$p < .001$.

best prediction of rate of attrition, and based on the multivariable analysis, accounted for the largest amount of unique variance. The results for the IM measures were similar, as all IM composite models were significant (*p*s <.01) and reflected the multivariable analyses: Basic and complex IM had smaller hazard ratios, Exp(β) = 1.004, and 1.001, compared to the motor and interference measures, Exp(β) = 1.043, and 1.012, which also significantly declined over time in study (β = -.015, and -.006). Based on the odds ratios, motor IM was the strongest indicator of rate of attrition. Therefore, the strongest predictor of rate of attrition was ISD based on moderately complex RT tasks (basic), but the complexity of the RT tasks did not appear to play a role for the IM measures, as the simplest task offered the best prediction.

Inconsistency and Neuropsychological Tests Predicting Rate of Attrition

Because basic ISD was the strongest ISD composite in predicting rate of attrition, it was used the comparison for the neuropsychological measures. We conducted a series of hierarchical regressions comparing the ability of basic ISD and its interaction with time to each neuropsychological task and their significant time interactions in predicting rate of attrition. The results are presented in Table 15, and were surprisingly consistent. MMSE and both WAIS subscales significantly accounted for unique variance over and above the effects of basic ISD and its time interaction. Basic ISD also accounted for unique variance over and above each of these measures, but the unique variance was significantly smaller. This indicates that although the neuropsychological tests were superior predictors of rate of attrition, basic ISD provided additional information that was not found by relying on the neuropsychological measures alone. On the other hand, the analyses involving word recall and Trails-B supported the opposite conclusion: Basic ISD accounted for a larger proportion of unique variance over and above word recall and

Table 15

Hierarchical Regressions of Basic ISD and Neuropsychological Tests Predicting Rate of Attrition

	Unique to N.psych Test	Unique to Basic ISD
Neuropsychological Test	χ^2 change	χ^2 change
MMSE	240.31***	13.42**
Word Recall	82.63***	98.00***
TMT – B	13.97**	129.15***
WAIS Block	94.12***	78.27***
WAIS Vocabulary	118.20***	35.72***

Notes. Each neuropsychological measure was individually compared with basic ISD. χ^2 change df = 2. MMSE = Mini-Mental State Examination; TMT-B = Trail Making Test – Part B. *$p < .05$; **$p < .01$; ***$p < .001$.

Trails-B than did word recall and Trails-B account for over and above the ISD composite. Overall, both the neuropsychological measures and basic ISD were uniquely important in predicting rate of attrition over 5 years, but MMSE and the WAIS block and vocabulary subscales accounted for a larger percentage of unique variance than ISD. Therefore, although basic ISD was not a superior predictor compared to the majority of the neuropsychological tasks, it did significantly provide additional information in predicting the rate of dropping out of the study.

Discussion

Through a series of 5-year prospective investigations, the present study evaluated the validity of the hypothesis that intraindividual variability is indicative of impending cognitive decline and neurological impairment. These investigations analyzed whether baseline intraindividual variability could significantly predict a range of meaningful 5-year change outcomes, including cognitive change, change in cognitive status classification, and sample attrition, and compared this prediction to those attained by mean level of performance. Further, potential differences in the predictive relationship due to task complexity were evaluated. This study was also the first to address whether intraindividual variability was a better predictor of later cognitive ability than standard neuropsychological predictors.

Cognitive Level

Together with information on age group at Wave 1, the set of ISD composites at baseline significantly predicted individuals' level of cognitive ability 5 years later. These results extended those found in Study 1 demonstrating a short-term 3-year link between inconsistency and cognition, and are consistent with even longer prediction intervals (6 years, MacDonald et al., 2003). Even with the addition of age group as a predictor, the

percent of variance overlap between the variability measures and the cognitive scores 5 years later was impressively similar to that reported across only 3 years in Study 1. In fact, correspondence between the two time intervals was found in every aspect of the predictive relationships: both studies found the group of predictors accounted for the largest proportions of variance for the digit symbol, letter series, and word recall tasks, and relatively smaller, but significant amounts for the vocabulary and similarities scores. In addition, the complex ISD composite provided the strongest unique prediction of later cognitive ability both 3 and 5 years later, and for both studies nearly all of the relationships between the ISD composites and the cognitive scores were in the expected negative direction (i.e., higher ISD, lower cognitive score). These common outcomes indicate that the nature of the predictive relationship between inconsistency and later cognitive score remained stable over time, but it is reasonable to expect that the strength of this relationship would diminish over extended time intervals. Explicit examinations comparing this possibility have yet to be completed. Regardless, the present results demonstrate the significant link between initial intraindividual variability and later cognitive score, and support the hypothesis that intraindividual variability may be an early marker of change in cognitive ability (e.g., Lövdén et al., 2007).

Cognitive Change

Although the ability to predict an individual's later cognitive level from their amount of inconsistency 5 years prior was significant, this information is not particularly meaningful to clinicians. As evidenced from the significant random effects found in Study 1, individuals vary in their level of cognitive ability, and as such it has been suggested that an individual's change over time, rather than their final level of performance, may be a superior indicator of the initial stages of non-normative aging

(e.g., Tuokko & Hultsch, 2006). Therefore, we also investigated whether inconsistency could predict meaningful 5-year change in each cognitive domain; did individuals show later decline, improvement, or stability based on their initial variability score?

Despite being able to significantly predict level of cognitive ability 5 years later, the set of ISD composites were not as good at predicting the cognitive change groups. For digit symbol, letter series, similarities, and vocabulary, the group of ISD predictors and age group at Wave 1 did not significantly distinguish among the various change groups. There were a handful of unique effects of the various predictors, but these only approached significance. Individual models were similarly poor predictors, showing the motor and interference ISD approached significance in distinguishing the digit symbol groups, and basic ISD was nearly significant for the vocabulary change groups. Overall, inconsistency was a poor predictor of change group on these tasks.

The results were more positive for word recall change groups however. The model test of all four ISD composites and age group at Wave 1 was significant, demonstrating that the group of predictors significantly differentiated between the improve, decline, and stable word recall change groups, where greater inconsistency was associated with a poorer change group. Both basic and complex ISD were particularly important to the model and despite having only marginally significant unique effects, the resulting odds ratios were impressive. For example, a one standard deviation increase in basic ISD at Wave 1 resulted in a 318% less likelihood of being in the stable group than the decline group, and a 347% less likelihood of being in the improve group than the decline group. The individual models revealed the complex, basic, and interference composites of inconsistency were all significant predictors, and motor ISD approached

significance. Although both the basic and complex composites had identical effect sizes, changes in basic ISD yielded stronger odds comparisons.

Generally speaking, inconsistency was a poor determinant of whether an individual showed reliable decline, improvement, or stability on a cognitive task over 5 years. However, this negative outcome was limited to the perceptual speed, reasoning, verbal fluency, and vocabulary domains. On the other hand, greater baseline inconsistency was particularly attuned to later changes in episodic memory. These findings corroborate earlier results showing intraindividual variability differentially predicted later cognitive ability, and showed stronger prediction and coupling of the fluid processing domains rather than crystallized knowledge (i.e., Study 1; MacDonald et al., 2003). However, Study 1 found the strongest annual coupling of change in inconsistency and change in cognition for perceptual speed, closely followed by episodic memory and reasoning. Therefore, it was unexpected for episodic memory to show the strongest group prediction by inconsistency, and for the other two fluid domains (i.e., perceptual speed and reasoning) to be so poorly predicted. This finding requires replication, but it is clear that intraindividual variability is particularly revealing for change in performing fluid-based cognitive tasks.

The limited links in predicting the 5-year cognitive change groups may also be due to the relatively smaller numbers of participants who showed change on the cognitive tasks. However, the likelihood of this having a meaningful effect is slim because significant differences were found for the word recall task, which showed similar cell sizes (see Table 6). Rather, stronger prediction might be found in a more impaired sample. As noted in Study 1, the present sample was relatively healthy and well-educated, and represented a more select group of older adults than might be randomly

found in the population. Further, because we required data on the participants after 5 years, only those participants who completed all 6 waves of testing were included in the present analyses. Individuals who remained in the sample were significantly younger (*M* = 73.41; Non-returning, *M* = 75.42), had fewer chronic conditions (*M* = 2.73; Non-returning, *M* = 3.37), and viewed themselves to be in better health than others their own age (*M* = 4.36; Non-returning, *M* = 3.95). Thus, stronger prediction of the change in other cognitive domains in a more diverse older sample looks promising.

The results for word recall also showed that the complexity of the inconsistency-based tasks are important in being able to later predict whether an individual will show reliable cognitive decline, improvement, or stability. However it was moderately complex tasks (i.e., basic ISD derived from CRT-type tasks) rather than the most demanding ones that showed the strongest ability to the differentiate 5-year word recall change groups. The superiority of this type of inconsistency for episodic memory is consistent with Study 1, which found inconsistency based on moderately challenging cognitive tasks showed similar coupling links to cognition as highly challenging cognitive tasks.

MMSE Change

Because the MMSE is a measure of overall cognitive functioning, significant change on this test can be a reliable sign of cognitive decline. We hypothesized that if inconsistency is sensitive to neurological disturbance, it may also be able to predict potential decline on MMSE even before it occurs. However, such was not the case. The various Wave 1 ISD composites and age group failed to significantly differentiate among the three MMSE 5-year change groups (improve, stable, decline), and further distinctions included interference ISD only approaching significance. Although the individual models

were non-significant, complex ISD could significantly distinguish the decline and stable groups: a one standard deviation increase in ISD on the complex RT tasks at Wave 1 reduced the likelihood of being in the stable MMSE change group 5 years later compared to being in the decline MMSE change group by 168%.

Despite cross-sectional findings that intraindividual variability can predict even subtle changes in pathological cognitive decline (e.g., CIND, Strauss et al., 2007), intraindividual variability was not a good predictor of change on the MMSE. There are several components that explain this lack of significant findings. First, the MMSE is a very general cognitive measure, where substantial declines in functioning are needed before they are reflected in the test score. The select and healthy nature of the sample meant they simply did not show substantial decline over the 5 years. In fact, over 80% of the sample showed a MSME score of 26 or higher out of a possible total score of 30. This relative stability was reflected in the cell sizes for the MMSE change groups, where only 17 participants showed the minimum clinically meaningful 3-point decline. Based on these results, inconsistency may be better suited to predicting change on specific cognitive tasks that are sensitive to subtle changes in ability. In fact, predicting decline in specific cognitive domains is likely more valuable in identifying preclinical impairments.

The finding that the most cognitively complex measure of inconsistency showed the only significant group comparison fits with the finding that little cognitive deterioration was evident in the present sample. In other words, the most difficult index of inconsistency appeared to the only measure that was able to find significant differences in the highly functioning sample.

Cognitive Status Change

Our hypothesis that baseline inconsistency may be able to identify those individuals with reliable CIND classifications (i.e., those who maintain CIND status or become CIND over time) was supported. The model with all Wave 1 ISD composites and age group was able to significantly predict CIND change group status (i.e., stable intact, fluctuate, stable decline, stable CIND). Complex ISD was particularly important to the model: for each standard deviation increase in complex ISD, an individual was 329% less likely to be in the stable intact group, and 290% less likely to be in the stable decline group, compared to the stable CIND group. Further, an individual was 186% less likely to be in the stable intact group compared to the fluctuating group. Individually, the basic, complex, and interference ISD models were significant, and although the complex model had the largest effect size, the basic ISD model had stronger odds ratios in distinguishing the groups. Specifically, basic ISD was able to significantly distinguish the stable intact group from the 3 other groups: A one point increase in basic ISD score at Wave 1 increased an individual's likelihood of being in the fluctuating group by 290%, the stable decline group by 609%, and the stable CIND group by 398%.

Clearly, inconsistency was a significant predictor of change in CIND status over time. These results reinforce the likelihood that neurological processes are the cause of moment-to-moment fluctuations in cognitive performance, and demonstrate that intraindividual variability may be a valuable tool in predicting preclinical dementia. This finding is particularly noteworthy because at the moment, clinicians are unable to reliably classify individuals as CIND (e.g., Tuokko & McDowell, 2006). However, because initial inconsistency did not significantly distinguish each group from each other group, the predictive power of inconsistency must be tempered. There were only three significant distinctions among the 4 change groups for even the strongest ISD composite

(i.e., basic ISD). Rather than this indicating unreliability of the link between inconsistency and later CIND status however, this more likely represents difficulties in the CIND change groups. For example, the CIND change groups were expected to follow a continuum of severity in cognitive impairment (i.e., stable intact < fluctuating < stable decline < stable CIND). However, the largest implication for a one unit increase in basic ISD at Wave 1 was an increased likelihood of being in the stable decline group (i.e., becoming CIND over the 5 years), rather than staying as stable intact (i.e., being intact or healthy over the 5 years). Rather, we would have expected the largest difference to exist between the two ends of the continuum: stable intact and stable CIND. The distinction between these two groups also significantly increased with each increase in basic ISD, but to a smaller extent than with the stable decline group. Another unexpected finding was that greater inconsistency at baseline did not differentiate among the three somewhat "impaired" change groups. Consequently, other methods of classifying these "impaired" change groups may be more appropriate. However, the finding that initial inconsistency significantly differentiated each of these groups from the one "healthy" group (i.e., stable intact), verifies the predictive utility of intraindividual variability over longer time periods. It remains to be seen whether all of the CIND change groups go on to develop dementia however, before actual clinical usage of inconsistency would be recommended.

Another interesting finding was the significant odds ratio found between the stable intact and fluctuating groups. The fluctuating group included individuals who changed from intact to CIND and back to intact (or vice versa) at least once over the 4 time points. Given the poor stability of CIND status, fluctuation in cognitive status classification was not unlikely, and it was uncertain whether this type of pattern reflected anything more than individuals having random fluctuations in their performance. Clearly, instability in

cognitive status was also meaningful, and potentially indicative of the initial stages of neurological disturbance.

Consistent with the results showing prediction of the other group outcomes, the complexity of the inconsistency-based tasks enhanced differentiation among the CIND change groups. However, the moderately complex RT tasks provided the best insight into predicting later cognitive change, with the most complex tasks showing similar results.

Attrition

A Cox regression model including all ISD composites, age group, and the associated significant interactions with time was indeed able to significantly differentiate those who dropped out of the study from those who remained in the study. For example, for a one standard deviation increase in motor ISD at Wave 1, the risk of attriting uniquely increased by 340%. Given these impressive findings, the considerable evidence showing links between inconsistency and various neurological problems (e.g., traumatic brain injury, Stuss et al., 2003; dementia, Hultsch et al., 2000; Parkinson's disease, Burton et al., 2006) and brain characteristics (e.g., corpus callosum size, Anstey et al., 2007; brain activation, Bellgrove et al., 2004; regulation of competing neural processes, Kelly et al., 2008), and the findings that all maladaptive change patterns of cognitive status in the present study showed initially higher inconsistency, attrition did appear to be a reasonable proxy for impending health problems. If participants were truly dropping out of the study for normative reasons such as lack of interest or time constraints, we would not have found such a strong prediction of attrition by baseline inconsistency.

Table 16 summarizes the self-reported reasons for participant drop-out at each wave. Although a number of participants provided normative justification for not continuing in the study after the first testing wave, the proportion of nonnormative

Table 16

Reported Reasons for Attrition

Wave	Returned	Attrited	
		Normative	Nonnormative
1	304	-	-
2	270*	25	9
3	256	7	7
4	239	7	10
5	234	1	4
6	218	11	5

Notes. Normative = Busy or not interested, family health problems, moved, other; Nonnormative = Died, memory or health problems. *Includes 16 participants who were not asked to return for Wave 2 testing due to budgetary restrictions, but were invited to participate in Wave 3.

reasons increased across the testing waves. To further investigate the possibility of a discrepancy between participants' reported reasons for drop-out and potential actual explanations, we drew a random sample of 10 of the 25 participants who dropped out for normative reasons after Wave 1. Although these participants described themselves as busy or not interested in further participating, 5 of the participants demonstrated poorer cognitive ability than their peers (i.e., CIND-Multiple based on their Wave 1 cognitive test scores), 5 participants had three or more chronic health conditions, and 4 participants completed fewer than 13 years of education[13]. Therefore, some of the participants who reported normative drop-out had potentially poorer health and cognitive abilities. Together with the strong demonstrated relationship between attrition and baseline inconsistency, this data supports the possibility that not all participants reported the true reasons for removing themselves from the study.

The present findings are consistent with those by MacDonald and colleagues (2003) which showed individuals who later dropped out a 6-year longitudinal study showed greater fluctuations in their cognitive performance at baseline. They also fit the results from studies investigating one particular reason for drop-out, death. Eizenman et al. (1997) found those who were more inconsistent in their control beliefs at the initial time point were more likely to die 5.5 years later. Further, the present results corroborate findings from an in press study by MacDonald and colleagues which used survival analysis to investigate time to death. They found inconsistency significantly increased per additional year closer to death, and that baseline intraindividual variability could predict impending death up to 15 years later. Clearly, intraindividual variability is a valid

[13] These 3 characteristics were not always found among the same participants. For example, 2 of the participants with poorer health had high education levels and intact CIND status.

early indicator of maladaptive outcomes, particularly attriting from a study and death, supporting the hypothesis that the inconsistency is a trait-like characteristic that reflects neurological integrity. Consequently, the ability to predict dementia via inconsistency is promising.

Not all tasks uniquely influenced predicting attrition rate, but these results were not consistent with the complexity hypothesis. In the overall model, complex ISD was surprisingly non-significant, as was interference ISD. On the other hand, participants who showed greater inconsistency on the motor and basic ISD tasks were at a substantially greater risk of dropping out of the study. For example, with each one SD increase in inconsistency on the basic RT tasks, the rate of attriting uniquely increased by an impressive 1605%. Individual models showed all of the ISD composites were significant predictors, but the results corresponded to those reported in the multivariable analyses showing motor and basic ISD had much larger hazard ratios. The significance of motor ISD was unexpected given its non-significant contributions in predicting the other change groupings (i.e., cognitive change, MMSE change, cognitive status change), but fluctuations on the simplest RT tasks (i.e., repetitive finger tapping) may be more reflective of general health that is more likely to be assessed by attrition than the other outcome measures. However, the superiority of the basic ISD measure in predicting rate of attrition is consistent with earlier findings showing moderately complex tasks appear to be most useful in prospective longitudinal relationships.

Interestingly, the added risk of being more inconsistent on the RT tasks slightly decreased with each additional year of being in the study. For example, the hazard ratio associated with showing greater inconsistency at Wave 1 was higher than that associated with subsequent waves of testing, and this effect was greater for performance on the

moderately complex tasks (i.e., Motor ISD X Time, β = -.116; Basic ISD X Time, β = -.278). Therefore, it appears that the initial amount of inconsistency was particularly informative on rate of attrition. This finding fits with the survival function for remaining in the study across the testing waves (see Figure 5). Cumulative survival dropped more between Waves 1 and 2 (i.e., between 0 and 365 days) than the later waves of testing, indicating that most participants who did not continue in the study chose to do so early on. It remains to be seen whether this effect is specific to the particular time interval used (i.e., 5 years), or whether a different mechanism is at work in the first year of a longitudinal study. For example, even those with slight cognitive impairment or health problems may be swept up in the excitement of being part of a new research study and volunteer to participate, thus influencing the links between initial intraindividual variability and the rate of drop-out.

Across the Various Outcomes

Interestingly, only specific 5-year outcomes were significantly predicted by baseline intraindividual variability. This included cognitive score for all five cognitive domains, word recall change group, cognitive status change group, and rate of attrition. An individual's likelihood of being in a poorer outcome group for these measures was impressively high with each unit increase in variability. The significant relationships with these particular outcomes, rather than the other domains of cognitive change or change on the MMSE, may be due to the fact that these outcomes are associated with the early behavioural characteristics of dementia, which clearly results from neurological disturbance. Episodic memory decline is one of the hallmark characteristics of early dementia (e.g., Albert, 2008; Bäckman et al., 2005), CIND status is based on potentially initial symptoms of dementia, and attrition from longitudinal studies is similarly believed

to be indicative of underlying influences such as disease and cognitive impairment (e.g., MacDonald et al., 2003; Sliwinski et al., 2003). Although the MMSE is a neuropsychological test, it is a measure of general cognitive functioning, and given the select nature of the sample, likely did not show the subtle changes in functioning that were apparent in the other outcome measures. The lack of findings for the other cognitive domains are also reasonable given that episodic memory is generally the first and most salient symptom of early dementia, but problems with reasoning, perceptual speed, and verbal ability tend not to appear until later in the disease process (Albert, 2008). Further, despite significant prediction of the level of cognitive ability 5 years later on all cognitive domains, significantly distinguishing cognitive change was clearly much more difficult in this healthy sample, and thus only significant for the measure with the greatest links to potential pathological decline. Consequently, the link between intraindividual variability and these specific deleterious outcomes corroborates hypotheses that fluctuations in behavioural performance are caused by neurological mechanisms (e.g., Li & Lindenberger, 1999).

Despite the remarkable strength between an individual's initial variability in responding and their cognitive change over time, age group was also a reliable predictor. In fact, age group at Wave 1 was a stronger predictor of many of the group outcomes than the inconsistency measures, including digit symbol, word recall, MMSE, and CIND change group (it is important to reiterate however, that inconsistency was a significant predictor of the various outcomes over and above this age group effect). In each case, being in the old-old age group (i.e., 75-92 years) at the initial wave of testing resulted in a greater risk of showing cognitive decline and deterioration 5 years later. The greater risk associated with being in the old-old age group was particularly striking in the survival

function of attrition (see Figure 6). For example, individuals who were in the old-old age group had a 15.55 times greater rate of attrition than those in the young-old age group. However, this enhanced risk decreased over time in study (i.e., Age Group X Time, β = -.859), supporting our hypothesis that the initial year of testing may recruit participants not well suited for longitudinal studies. The substantial influence of age group on predicting 5-year change outcomes was not unexpected given the greater risk of decline and disease with older age, and reiterates the importance of including biological age in estimating prospective outcomes.

Intraindividual Variability Versus Intraindividual Mean

Generally speaking, intraindividual variability and intraindividual mean were comparable in differentiating among the various change outcomes. For outcomes where the ISD models were poor predictors of change group, the IM models were also not significant (i.e., digit symbol, letter series, similarities, vocabulary, MMSE change). Similarly, for outcomes where inconsistency significantly distinguished among the change groups, mean rate of responding did so as well (e.g., word recall change, cognitive status change). Further, the two measures often had similar effect sizes in predicting the change groups, such as for word recall change (ISD: Nagelkerke's R^2 = .18; IM: Nagelkerke's R^2 = .17), and the significant unique effects for corresponding task composites (i.e., basic ISD, basic IM).

However, there were some distinctions between inconsistency and mean level of responding. The superiority of the two measures varied based on which cognitive domain task was being predicted 5 years later. The ISD measures added unique information after the entry of the IM measures for vocabulary, but the IM measures did not significantly add to the prediction of the word recall, similarities, and vocabulary scores after entry of

the ISD measures. On the other hand, the IM measures appeared to be a better predictor of later digit symbol score than the ISD measures, and both of the measures provided significant information over and above one another for letter series. Thus, there did not appear to be a reasonable overall pattern. The differences were similarly mixed for predicting the cognitive status change groups. On the one hand, there were fewer significant group distinctions by the individual IM models for some composites (see Table 12). For example, basic ISD significantly differentiated the stable intact group from the 3 other groups (overall Nagelkerke's R^2 = .08), but the basic IM model was only marginally significant (Nagelkerke's R^2 = .04) and only distinguished the stable intact and stable CIND groups. On the other hand, neither measure significantly distinguished the groups with the motor composite, and each showed the same significant group differentiations for the complex composite. Comparable differences between inconsistency and the mean were also found in predicting rate of attrition, where the two models were nearly equivalent, but differed in which particular composites were significant. The basic and motor composites were the strongest ISD predictors, while the interference and motor composites were the strongest IM predictors.

The evaluations between intraindividual variability and the mean were also hampered by the difficulty in comparing the resulting odds ratios for the two measures directly. It would have been insightful to see whether a one unit increase in inconsistency had a greater or lesser impact on the odds of being in one group, compared to a one unit increase in average speed of responding. For example, a one unit increase in basic ISD score at Wave 1 increased an individual's likelihood of being in the stable CIND group rather than the stable intact group by 38%, whereas a comparable increase in basic IM increased the likelihood of by only 23%. However, we were unable to perform

these contrasts because of the different scale of the two measures: A unit increase in ISD was one tenth of a standard deviation in variability, whereas a one unit increase in IM was one millisecond of change in reaction time. Consequently, we were unable to say whether an individual was at greater risk of being in a maladaptive outcome group if they were more inconsistent, or they were slower in responding.

Overall, the present results do not appear to easily fit with previous research comparing the two measures (e.g., Bielak, Hughes et al., 2007; Christensen et al., 2005; Hultsch et al., 2000). There are some instances of the mean and inconsistency showing identical predictive ability of group outcomes (e.g., word recall change, MMSE change), but also examples where inconsistency did offer information that differed from that provided by the mean level of performance (e.g., cognitive status change). The differences tended to preside across the individual composites rather than the overall models however (e.g., the strongest predictors of rate of attrition for mean and inconsistency were motor IM and basic ISD respectively). Therefore, is it the case that short-term variability does in fact not offer any unique predictive power beyond mean level of performance when comparing performance on some measures, but is particularly informative on others? Further research on this topic is needed, but it should not overshadow the substantial evidence already found between intraindividual variability and various characteristics and conditions (e.g., corpus callosum size, Anstey et al., 2007; Parkinson's disease, Burton et al., 2006; mild cognitive impairment, Dixon et al., 2007; older age, Hultsch et al., 2002; poorer cognitive performance, Li et al., 2001; availability of neurotransmitters, Li & Lindenberger, 1999; physical ability, Strauss et al., 2002) that point towards the likelihood that inconsistency is indicative of neurological integrity. In addition, recent findings showing the significant prediction of various maladaptive

outcomes by inconsistency (e.g., MacDonald et al., in press), and that inconsistency, but not mean performance, reliably preceded and predicted subsequent decline in cognition (Lövdén et al., 2007), are particularly promising.

Differences Due to Task Complexity

There were clear differences in the predictive relationships based on the complexity of the RT tasks. For inconsistency, the most cognitively demanding tasks were strongest at predicting level of cognitive ability and MMSE change group. For the other outcomes of word recall change group, cognitive status group, and rate of attrition however, moderately challenging tasks provided the most insight. It appears that intraindividual variability that is derived from RT tasks that provide moderate to high cognitive challenge may be the most sensitive to longitudinal changes in cognitive ability. These results are consistent with findings that inconsistency based on cognitively demanding tasks provided greater sensitivity to various outcomes (e.g., mild cognitive impairment, Strauss et al., 2007). For the mean level of performance, the complexity of the RT tasks did not play as clear a role. Although complex IM was the strongest predictor of later cognitive level and cognitive status change group, the best predictor of the rest of the change outcomes varied: basic IM for word recall change, motor IM for rate of attrition, and no composites were significant for MMSE change. These results suggest particular attention should be given to the type of RT tasks used to calculate inconsistency, and because no clear pattern has emerged, a wide range of RT tasks should be used as the basis for mean rate of responding.

Intraindividual Variability versus Neuropsychological Tests in Predicting Attrition

Although every neuropsychological test significantly accounted for unique variance over and above the effects of basic ISD, basic ISD also provided additional

information that was not found by relying on the neuropsychological measures alone. However, the amount of unique information offered by the initial score on the MMSE, WAIS block, and WAIS vocabulary subscales was substantially greater than that offered by basic ISD. On the other hand, basic ISD accounted for a larger percentage of unique variance than word recall and Trails-B. Therefore, it appears that inconsistency may be more attuned to the nonnormative reasons for attriting from the study (i.e., deteriorating health, abnormal aging and disease, or impending death) than clinical measures assessing episodic memory and executive functioning. Interestingly, tests of more general and crystallized functioning appear to be better predictors of rate of attrition than intraindividual variability. Given its demonstrated links to neurological processes (e.g., Kelly et al., 2008), it may be the case that intraindividual variability is an early indicator of potential drop-out for nonnormative reasons, and that general tests of functioning and those that are stable with age (i.e., vocabulary) are markers of normative reasons for attriting (e.g., time constraints and lack of interest. The comparison between WAIS block and basic ISD does not neatly fit this hypothesis, but the idea is intriguing nonetheless. Overall, both the neuropsychological measures and basic ISD were uniquely important in predicting rate of attrition over 5 years, and basic ISD offered a significant amount of unique information in every comparison. At this point in time, there is insufficient evidence to support clinicians including intraindividual variability measures as part of their diagnostic inventory, but this first evaluation in predicting the rate of attrition is promising. Future research predicting impending dementia will be particularly enlightening on this issue.

Limitations and Future Directions

As noted earlier, the present sample was relatively healthier, better educated, and thus more selective than the general population. Further, there was evidence of selective attrition from the study, in terms of those dropping out of the study being older and in poorer health than those who maintained participation over the 5 years. Both of these factors were evident in the little change that occurred over the 5 years on the various cognitive measures, and the scarcity of dementia diagnoses. Consequently, the fact that the present significant results were found even in such a healthy sample demonstrates they are particularly robust, and the effects of prediction are likely to be even higher in a more diverse older adult sample.

Another limitation of the present study was the creation of the CIND change groups. As noted earlier, the lack of differentiation among the three maladaptive groups suggests superior groupings may exist. For example, we did not differentiate among those who minimally fluctuated between CIND and intact status versus those who fluctuated a great deal (i.e., CIND, intact, CIND, intact), and we did not take into account the time point at which individuals appeared to remain stable in their CIND classification (e.g., intact, intact, CIND, CIND, versus intact, CIND, CIND, CIND). There were numerous other change group distinctions that could have been made but given the size of the sample, the four change groups were reasonable and well-justified choices (see methods section).

Finally, there is little chance that all individuals who dropped out of the study did so because of declining health or cognitive ability. Rather, many participants likely became too busy to participate for 7 sessions every year, or simply did not enjoy the testing environment or the tests themselves. However, based on the present results and

those of others (MacDonald et al., 2003; Sliwinski et al., 2003), attrition appears to be a reasonable and reliable proxy for disease and impairment in the older adult population.

Overall, the present study showed strong support for the sensitivity of intraindividual variability to cognitive change over 5 years. The initial level of inconsistency was particularly sensitive to changes associated with the early behavioural characteristics of dementia, including episodic memory ability, cognitive status, and attrition. In each case, greater inconsistency at baseline was associated with a greater likelihood of being in a maladaptive group 5 years later. Mean rate of responding was a comparable predictor of change in most instances, but differences emerged according to the complexity of their derived tasks. Variability based on moderate to high cognitively challenging tasks appeared to be the most sensitive to longitudinal changes in cognitive ability, but no clear pattern was evident for mean speed of responding. Although inconsistency was not a superior predictor compared to the majority of the neuropsychological tasks, it did significantly provide additional information in predicting the rate of attrition. The potential utility of intraindividual variability as an early marker of neurological disturbance and cognitive decline is promising, but further study is needed across longer time frames, and involving various neurological conditions.

General Discussion

This dissertation presented two studies of intraindividual variability in a longitudinal context to further explore the relationship between short-term intraindividual variability and longer-term cognitive change in older adults. The first study addressed the longitudinal nature of intraindividual variability over 3 years. On average, those aged 75 and older showed substantial increases in inconsistency over time, but there were significant individual differences in change across the entire sample (i.e., 64-92 years).

The significant covariation relationship between change in cognition and change in inconsistency was replicated and extended to intervals as short as one year, and found to remain stable across both time and older age. The strength of the coupling relationship however did differ based on the cognitive difficulty of both the cognitive domain and the inconsistency-based tasks.

Building on the results of the first study suggesting that intraindividual variability is highly sensitive to even subtle changes in cognitive ability, the second study addressed the capacity of intraindividual variability to predict cognitive ability and other meaningful change outcomes 5 years later. Inconsistency at Wave 1 was particularly sensitive to changes associated with the early behavioural characteristics of dementia, including episodic memory ability, cognitive status, and attrition, but intraindividual mean was a comparable predictor of change. Variability based on moderate to high cognitively challenging tasks appeared to be the most sensitive to longitudinal changes in cognitive ability, and was uniquely predictive of the rate of attrition compared to neuropsychological tasks.

The present studies expanded the literature on intraindividual variability in two important ways: Theoretically in Study 1 by reiterating the trait-like nature of intraindividual variability as the negative coupling relationship between cognition and inconsistency was maintained over annual intervals; and applicably in Study 2 by showing intraindividual variability truly is an early marker of cognitive decline and longitudinal outcomes. Consistent with theoretical suggestions by Nesselroade (1991) and Hultsch and colleagues (2008), this pair of studies demonstrated that changes in the distributions of short-term variability are indeed linked to and indicative of longitudinal intraindividual cognitive change. Further, the almost exclusive direction of the coupling

(i.e., higher ISD, poorer cognition) and predictive relationships (i.e., each unit increase in ISD resulted in a greater likelihood of being in maladaptive group 5 years later), and the specific predictive sensitivity to neurologically-linked outcomes supports the hypothesis that greater inconsistency is a behavioural marker of neurological disturbance (Li & Lindenberger, 1999).

There was also enough similarity between the research questions addressed in the two studies that replication and extension of certain results was feasible. For example, there was impressive correspondence between the prediction of cognitive ability at the two time intervals (i.e., 3 years and 5 years): Both studies found baseline inconsistency accounted for the largest proportions of variance for the digit symbol, letter series, and word recall tasks, and that ISD based on particularly challenging RT tasks provided the strongest prediction. Evidence from the annual coupling relationships confirmed the sensitivity between fluid cognitive domains such as memory, reasoning, and processing speed, and inconsistency in performance, but also showed that fluctuations on moderately complex tasks (i.e., basic ISD) can be equally good indicators of change in certain cognitive domains. These results were enhanced even further from the predictions of meaningful 5-year change outcomes, which showed word recall was the only cognitive domain whose change groups were significantly predicted. However, inconsistency based on moderately complex RT tasks (i.e., basic ISD) rather than the most demanding ones (i.e., complex ISD) appeared to be the most useful in prospective longitudinal relationships. Consistent with previous findings (e.g., MacDonald et al., 2003; Strauss et al., 2007) there appears to be a clear trend for a stronger relationship between fluid types of cognition and inconsistency on more challenging measures.

Given the recent significant findings from MacDonald and colleagues (2003; in press), Lövdén and colleagues (2007), and the present body of work, researchers can infer with some degree of confidence that the longitudinal link between inconsistency and cognition is legitimate. Further, with the promising findings that inconsistency is sensitive to changes in cognition and other meaningful change outcomes (Lövdén et al., 2007; Study 2), it is clear that the future of this field is in longitudinal and predictive relationships. Specifically, the timescales deserve additional focus, including both shorter intervals for the coupling relationship with cognition, and longer time periods of the predictive relationships. The temporal order between cognition and intraindividual variability requires replication, and the longitudinal and predictive cognitive relationships need to be extended to other age groups and populations. Of particular importance, given the hypothesis that intraindividual variability is an early indicator of behavioural deterioration, is investigating whether baseline variability can predict the diagnosis of a neurological condition such as dementia years later. Further, the prospective research needs be expanded to other neurological conditions, as inconsistency may be a stronger predictor of certain pathologies, particularly those affecting the frontal lobes (e.g., Stuss et al., 2003). Finally, greater attention needs to be focused on the mechanisms underlying fluctuations in performance by performing studies that directly link behavioural outcomes with brain-based processes (MacDonald et al., 2006), investigating more than one potential causative neurological process at a time, and eventually linking all three hypothesized processes showing causation from the brain to behavioural variability to cognitive change. Clearly, a plethora of research remains to be done in this area, but the recent findings are promising of the potential utility and applicability of intraindividual

variability in understanding and predicting intraindividual cognitive change in older adulthood.

References

Albert, M. S. (2008). The neuropsychology of the development of Alzheimer's disease. In F. I. M. Craik and T. A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed., pp. 97-132). New York: Psychology Press.

Allaire, J. C., & Marsiske, M. (2005). Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychology and Aging, 20*, 390-401.

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

Anstey, K. J., Dear, K., Christensen, H., & Jorm, A. F. (2005). Biomarkers, health, lifestyle, and demographic variables as correlates of reaction time performance in early, middle, and late adulthood. *The Quarterly Journal of Experimental Psychology*, *58A*, 5-21.

Anstey, K. J., Mack, H. A., Christensen, H., Li, S-C., Reglade-Meslin, C., Maller, J., et al. (2007). Corpus callosum size, reaction time speed and variability in mild cognitive disorders and in a normative sample. *Neuropsychologia*, *45*, 1911-1920.

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*, 520-531.

Baudic, S., Barba, G. D., Thibaudet, M. C., Smagghe, A., Remy, P., & Traykov, L. (2006). Executive function deficits in early Alzheimer's disease and their relations with episodic memory. *Archives of Clinical Neuropsychology*, *21*, 15-21.

Bellgrove, M. A., Hester, R., & Garavan, H. (2004). The functional neuroanatomical correlates of response variability: evidence from a response inhibition task. *Neuropsychologia*, *42*, 1910-1916.

Bielak, A. A. M., Hughes, T. F., Small, B. J., & Dixon, R. A. (2007). It's never too late to engage in lifestyle activities: Significant concurrent but not change relationships between lifestyle activities and cognitive speed. *Journals of Gerontology: Psychological Sciences, 62B,* P331-P339.

Bielak, A. A. M., Hultsch, D. F., Levy-Ajzenkopf, J., MacDonald, S. W. S., Hunter, M. A., & Strauss, E. (2007). Short-term changes in general and memory-specific control beliefs and their relationship to cognition in younger and older adults. *The International Journal of Aging & Human Development*, *65*, 53-71.

Bielak, A. A. M., Hultsch, D. F., Kadlec, H., & Strauss, E. (2007). Was that part of the story or did I just think so? Age and cognitive status differences in inference and story recognition. *Experimental Aging Research*, *33*, 295-322.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist*, *3*, 129-136.

Bleiberg, J., Garmoe, W. S., Halpern, E. L., Reeves, D. L., & Nadler, J. D. (1997). Consistency of within-day and across-day performance after mild brain injury. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *10*, 247-253.

Bunce, D., Anstey, K. J., Christensen, H., Dear, K., Wen, W., & Sachdev, P. (2007). White matter hyperintensities and within-person variability in community-dwelling adults aged 60-64 years. *Neuropsychologia*, *45*, 2009-2015.

Bunce, D., MacDonald, S. W. S., & Hultsch, D. F. (2004). Inconsistency in serial choice decision and motor reaction time dissociate in younger and older adults. *Brain and Cognition, 56*, 320-327.

Burton, C. L., Hultsch, D. F., Strauss, E., & Hunter, M. A. (2002). Intraindividual variability in physical and emotional functioning: Comparison of adults with traumatic brain injuries and healthy adults. *The Clinical Neuropsychologist*, *16*, 264-279.

Burton, C. L., Strauss, E., Hunter, M. A., & Hultsch, D. F. (in press). The relationship between everyday problem solving and inconsistency in reaction time in older adults. *Aging, Neuropsychology, and Cognition*.

Burton, C. L., Strauss, E., Hultsch, D. F., Moll, A., & Hunter, M. A. (2006). Intraindividual variability as a marker of neurological dysfunction: A comparison of Alzheimer's disease and Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology, 28*, 67-83.

Christensen, H., Dear, K. B. G., Anstey, K. J., Parslow, R. A., Sachdev, P., & Jorm, A. F. (2005). Within-occasion intraindividual variability and preclinical diagnostic status: Is intraindividual variability an indicator of mild cognitive impairment? *Neuropsychology, 19*, 309-317.

Clark, C. M., Sheppard, L., Fillenbaum, G. G., Galasko, D., Morris, J. D., Koss, E., et al. (1999). Variability in annual Mini-Mental State Examination score in patients with probable Alzheimer disease: A clinical perspective of data from the Consortium to Establish a Registry for Alzheimer Disease. *Archives of Neurology*, *56*, 857-862.

Collins, L. F., & Long, C. J. (1996). Visual reaction time and its relationship to neuropsychological test performance. *Archives of Clinical Neuropsychology*, *11*, 613-623.

Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging, Neuropsychology, and Cognition*, *12*, 187-215.

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. *Psychology and Aging*, *21*, 62-73.

Dixon, R. A., Garrett, D. D., Lentz, T. L., MacDonald, S. W. S., Strauss, E., & Hultsch, D. F. (2007). Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. *Neuropsychology*, 21, 381-399.

Eizenman, D. R., Nesselroade, J. R., Featherman, D. L., & Rowe, J. W. (1997). Intraindividual variability in perceived control in an older sample: The MacArthur successful aging studies. *Psychology and Aging*, *12*, 489-502.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Folstein, M., Folstein, S., & 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*, 189-198.

Fozard, J. L., Vercruyssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore Longitudinal Study of Aging. *Journal of Gerontology: Psychological Sciences*, *49*, P179-P189.

Fuentes, K., Hunter, M. A., Strauss, E., & Hultsch, D. F. (2001). Intraindividual variability in cognitive performance in persons with chronic fatigue syndrome. *The Clinical Neuropsychologist*, *15,* 210-227.

Hawley, D. R. (1995). Assessing change with preventive interventions: The reliable change index. *Family Relations*, *44*, 278-284.

Head, H. (1926). *Aphasia and kindred disorders of speech.* Cambridge, UK: Cambridge University Press.

Hetherington, C. R., Stuss, D. T., & Finlayson, M. A. J. (1996). Reaction time and variability 5 and 10 years after traumatic brain injury. *Brain Injury*, *10*, 473-486.

Hultsch, D. F., Hertzog, C., & Dixon, R. A. (1990). Ability correlates of memory performance in adulthood and aging. *Psychology and Aging, 5*, 356-368.

Hultsch, D. F., & MacDonald, S. W. S. (2004). Intraindividual variability in performance as a theoretical window onto cognitive aging. In R. A. Dixon, L. Bäckman, and L-G. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 65-88). New York: Oxford University Press.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology: Psychological Sciences, 57B*, P101-P115.

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology, 14,* 588-598.

Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. S. (2008). Intraindividual variability, cognition, and aging. In F. I. M. Craik and T. A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed., pp. 491-556). New York: Psychology Press.

Hunter, M. A., & Bielak, A. A. M. (2005). *Comparison of purification methods.* Unpublished manuscript.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12-19.

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93-132). Berlin: Springer-Verlag.

Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences*, *13*, 869-881.

Kelly, A. M. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2008). Competition between functional brain networks mediates behavioural variability. *NeuroImage*, *39*, 527-537.

Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *The Gerontologist, 9*, 179-186.

Li, S-C., Aggen, S. H., Nesselroade, J. R., & Baltes, P. B. (2001). Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance: The MacArthur successful aging studies. *Gerontology, 47*, 100-116.

Li, S-C., Huxhold, O., & Schmiedek, F. (2004). Aging and attenuated processing robustness. Evidence from cognitive and sensorimotor functioning. *Gerontology*, *50*, 28-34.

Li, S-C., & Lindenberger, U. (1999). Cross-level unification: A computational exploration of the link between neurotransmitter systems and dedifferentiation of cognitive abilities in old age. In L-G. Nilsson and H. J. Markowitsch (Eds.), *Cognitive neuroscience of memory* (pp. 103-146). Seattle: Hogrefe & Huber.

Li, S-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science*, *15*, 155-163.

Lövdén, M., Li, S-C., Shing, Y. L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. *Neuropsychologia*, *45*, 2827-2838.

MacDonald, S. W. S., Hultsch, D. F., & Bunce, D. (2006). Intraindividual variability in vigilance performance: Does degrading visual stimuli mimic age-related "neural noise"? *Journal of Clinical and Experimental Neuropsychology*, *28*, 655-675.

MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria Longitudinal Study. *Psychology and Aging, 18*, 510-523.

MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (in press). Intraindividual variability in neurocognitive performance predicts impending death. *Psychology and Aging*.

MacDonald, S. W. S., Nyberg, L. & Bäckman, L. (2006). Intra-individual variability in behaviour: links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences*, *29*, 474-480.

Marcos, A., Gil, P., Barabash, A., Rodriguez, R., Encinas, M., Fernandez, C., et al. (2006). Neuropsychological markers of progression from mild cognitive impairment to Alzheimer's disease. *American Journal of Alzheimer's Disease & Other Dementias*, *21*, 189-196.

Martin, M., & Hofer, S. (2004). Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology, 50*, 7-11.

Murtha, S., Cismaru, R., Waechter, R., & Chertkow, H. (2002). Increased variability accompanies frontal lobe damage in dementia. *Journal of the International Neuropsychological Society, 8,* 360-372.

Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In Roger M. Downs, Lynn S. Liben, & David S. Palermo (Eds.), *Visions of aesthetics, the environment & development: The legacy of Joachin F. Wohlwill* (pp. 214-242). Hillsdale, NJ: Erlbaum.

Nesselroade, J. R., & Featherman, D. L. (1997). Establishing a reference frame against which to chart age-related changes. In M. A. Hardy (Ed.), *Studying aging and social change: Conceptual and methodological issues* (pp. 191-205). Newbury Park, CA: Sage.

Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *Journal of Gerontology: Psychological Sciences, 59B*, P49-P55.

Peterson, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., & Kokman, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology, 56*, 303-308.

Psychological Corporation. (1997). *WAIS-III technical manual*. San Antonio, TX: Psychological Corporation.

Rabbitt, P., Osman, P., Moore, B., & Stollery, B. (2001). There are stable individual differences in performance variability, both from moment to moment and from day to day. *The Quarterly Journal of Experimental Psychology*, *54A*, 981-1003.

Ram, N., Rabbitt, P., Stollery, B., & Nesselroade, J. R. (2005). Cognitive performance inconsistency: Intraindividual change and variability. *Psychology and Aging, 20*, 623-633.

Reimers, S., & Maylor, E. A. (2006). Gender effects on reaction time variability and trial-to-trial performance: Reply to Deary and Der (2005). *Aging, Neuropsychology, and Cognition*, *13*, 479-489.

Reitan, R. (1958). Validity of the Trailmaking Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8,* 271–276.

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*, 140-144.

Salthouse, T. A., Nesselroade, J. R., & Berish, D. E. (2006). Short-term variability in cognitive performance and the calibration of longitudinal change. *Journal of Gerontology: Psychological Sciences, 61B*, P144-P151.

Salthouse, T. A., Toth, J. P., Hancock, H. E., & Woodard, J. L. (1997). Controlled and automatic forms of memory and attention: Process purity and the uniqueness of age-related influences. *Journal of Gerontology: Psychological Sciences, 52B*, P216-P228.

Sattler, J. M., & Ryan, J. J. (1999). *Assessment of children: Revised and updated third edition WAIS-III supplement.* La Mesa, CA: Sattler.

Schaie, K. W. (1989). Perceptual speed in adulthood: Cross-sectional and longitudinal studies. *Psychology and Aging*, *4*, 443–453.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, 1-5.

Sliwinski, M. J., Hofer, S. M., Hall, C., Buschke, H., & Lipton, R. B. (2003). Modeling memory decline in older adults: The importance of preclinical dementia, attrition, and chronological age. *Psychology and Aging*, *18*, 658-671.

Sliwinski, M. J., Smyth, J. M., Hofer, S. M., & Stawski, R. S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging*, *21*, 545-557.

Strauss, E., Bielak, A. A. M., Bunce, D., Hunter, M. A., & Hultsch, D. F. (2007). Within-person variability in response speed as an indicator of mild cognitive impairment in older adults. *Aging, Neuropsychology, and Cognition, 14,* 608-630.

Strauss, E., MacDonald, S. W. S., Hunter, M., Moll, A., & Hultsch, D. F. (2002). Intraindividual variability in cognitive performance in three groups of older adults: Cross-domain links to physical status and self-perceived affect and beliefs. *Journal of the International Neuropsychological Society*, *8*, 893-906.

Stuss, D. T., Murphy, K. J., Binns, M. A., & Alexander, M. P. (2003). Staying on the job: The frontal lobes control individual performance variability. *Brain*, *126*, 2363-2380.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Thurstone, T. G. (1962). *Primary mental abilities: Grades 9-12, 1962 revision.* Chicago: Science Research Associates.

Tuokko, H., & Frerichs, R. J. (2000). Cognitive impairment with no dementia (CIND): Longitudinal studies, the findings, and the issues. *The Clinical Neuropsychologist, 14*, 504-525.

Tuokko, H. A., Gabriel, G., & the CSHA Neuropsychology Working Group. (2006). Neuropsychological detection of cognitive impairment: Inter-rater agreement and factors affecting clinical decision-making. *Journal of the International Neuropsychological Society*, *12*, 72-79.

Tuokko, H. A., & Hultsch, D. F. (2006). The future of mild cognitive impairment. In H. A. Tuokko & D. F. Hultsch (Eds.), *Mild cognitive impairment: International perspectives* (pp. 291-304). New York: Taylor & Francis.

Tuokko, H. A., & McDowell, I. (2006). An overview of mild cognitive impairment. In H. A. Tuokko & D. F. Hultsch (Eds.), *Mild cognitive impairment: International perspectives* (pp. 3-28). New York: Taylor & Francis.

Walker, M. P., Ayre, G. A., Cummings, J. L., Wesnes, K., McKeith, I. G., O'Brien, J. T. et al. (2000). Quantifying fluctuation in dementia with Lewy bodies, Alzheimer's disease, and vascular dementia. *Neurology*, *54*, 1616-1624.

Wechsler, D. (1981). *Wechsler adult intelligence scale – Revised*. New York: Psychological Corporation.

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition, 49*, 402-419.

Williams, B. R., Hultsch, D. F., Strauss, E. H., Hunter, M. A, & Tannock, R. (2005). Inconsistency in reaction time across the life span. *Neuropsychology, 19*, 88-96.

Williams, B. R., Strauss, E. H., Hultsch, D. F., & Hunter, M. A. (2007). Reaction time inconsistency in a spatial stroop task: Age-related differences through childhood and adulthood. *Aging, Neuropsychology, and Cognition*, *14*, 417-439.

Wright, R. E. (2000). Survival analysis. In L. G. Grimm, & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 363-407). Washington, DC: American Psychological Association.