

Accepted in: Journal of the International Neuropsychological Society (Author copy)

A systematic review of longitudinal associations between reaction time intraindividual variability and age-related cognitive decline or impairment, dementia and mortality

Becky I. Haynes¹, Sarah Bauermeister² and David Bunce¹

¹ School of Psychology, University of Leeds

² Department of Psychiatry, University of Oxford

Address for correspondence: David Bunce, School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, LS2 9JT, UK. Email: d.bunce@leeds.ac.uk. Phone: +44(0)113 3435755

Abstract

Objective: Intraindividual variability (IIV) in reaction time refers to the trial-to-trial fluctuations in responding across a given cognitive task. Cross-sectional research suggests that IIV increases with normal and neuropathological ageing and it may serve as a marker of neurobiological integrity. This raises the possibility that IIV may also predict future cognitive decline and, indeed, neuropathology. Therefore, we conducted a systematic review to address these issues.

Methods: A search of electronic databases Embase, Medline, PsycINFO and Web of Science was completed on 17th May 2016 that identified longitudinal investigations of IIV in middle-aged or older adults.

Results: 688 studies were initially identified of which 22 met the inclusion criteria. Nine included longitudinal IIV measures and 17 predicted subsequent outcome (cognitive decline or impairment, dementia, mortality) from baseline IIV. The results suggested IIV increased over time, particularly in participants aged over 75 years. Greater baseline IIV was consistently associated with increased risk of adverse outcomes including cognitive decline or impairment, and mortality.

Conclusion: Increased IIV over time is associated with normal ageing. However, further increases in IIV over and above those found in normal ageing may be a risk factor for future cognitive impairment or mortality. Measures of IIV may therefore have considerable potential as a supplement to existing clinical assessment to aid identification of individuals at risk of adverse outcomes such as dementia or death.

Key words: Intraindividual variability; Ageing; Dementia; Cognitive decline; Mortality; Longitudinal.

It is well established that adult ageing is characterised behaviourally by increased intraindividual variability in cognitive function. As the present review will show, such variability is not only a feature of normal ageing, but becomes more marked in the presence of neuropathology or impending mortality. It is likely that this widely observed behavioural characteristic of ageing reflects greater neurobiological variability stemming from compromised central nervous system integrity. Given the proposed neurobiological underpinnings of IIV, the aim of the present systematic review, therefore, was to critically evaluate the empirical literature using this marker to predict cognitive change in normal ageing, and also age-related neuropathological outcomes including mild cognitive impairment (MCI), dementia, Parkinson's disease and, indeed, death. First though, we describe some of the characteristics of this behavioural marker before detailing theoretical and empirical linkage to potential underlying neurobiological variability.

Behavioural Intraindividual Variability

IIV refers to the within-person variation in cognitive performance. It is measured over relatively short periods using trial-to-trial fluctuations in reaction time (RT) on a given task, through repeated assessments over longer periods such as days or weeks, or across a battery of different cognitive tasks measured in the same session (Hultsch, MacDonald, & Dixon, 2002). The most widely used measure of IIV is RT variability, which provides the focus for the present review. Often, researchers have implicitly treated RT variability as random noise or error variance and collapsed trials across a cognitive task to obtain measures of central tendency (e.g., mean or median RT). However, theorists have suggested that this variation is systematic and that the measure may capture meaningful information about higher-order cognitive processes such as fluctuating attentional or executive control mechanisms (Bunce, MacDonald, & Hultsch, 2004; Bunce, Warr, & Cochrane, 1993; West, Murphy, Armilio, Craik, & Stuss, 2002). Moreover, research typically shows greater IIV to increase across the

adult lifespan (Dykiert, Der, Starr, & Deary, 2012; Hultsch et al., 2002; Williams, Hultsch, Strauss, Hunter, & Tannock, 2005). However, although it appears that IIV increases with normal ageing, due to the cross-sectional nature of many studies, the possibility of cohort effects cannot be ruled out.

Neuropathology and Brain Substrates of Intraindividual Variability

There is considerable empirical support for the suggestion that IIV reflects underlying neurobiological integrity (Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000; Hultsch, Strauss, Hunter, & MacDonald, 2008). For example, cross-sectional studies show that relative to healthy older persons, IIV is greater in individuals living with a range of neurodegenerative disorders including Parkinson's disease (e.g., de Frias, Dixon, Fisher, & Camicioli, 2007), MCI (e.g., Christensen et al., 2005; Dixon et al., 2007), and dementia (e.g., Gorus, De Raedt, Lambert, Lemper, & Mets, 2008; Hultsch et al., 2000), including early stage Alzheimer's disease (Duchek et al., 2009). However, given these findings of greater variability in the presence of age-related neuropathology, what evidence is there of associations between IIV and brain substrates? Several magnetic resonance imaging (MRI) studies have identified associations between greater behavioural variability and poorer neuroanatomical integrity as shown by reduced white matter volume (Jackson, Balota, Duchek, & Head, 2012; Walhovd & Fjell, 2007), increased white matter hyperintensity volume (Bunce et al., 2010; Bunce et al., 2007), and diffusion tensor imaging metrics such as fractional anisotropy (Deary et al., 2006; Fjell, Westlye, Amlie, & Walhovd, 2011; Mella, de Ribaupierre, Eagleson, & de Ribaupierre, 2013; Moy et al., 2011).

A plausible explanation for increases in IIV across the lifespan concerns the possibility that age-related dopamine depletion reduces neural signal-to-noise thereby affecting the efficiency of brain connectivity (Li, Lindenberger, & Sikstrom, 2001). Due to the hypothesized more intermittent signalling, variability increases, a possibility that has

received support from MRI work (e.g., MacDonald, Karlsson, Rieckmann, Nyberg, & Backman, 2012). Additionally, with the presence of amyloid, neurofibrillary tangles and the deterioration of white and grey matter structures that accompany the advance of age-related neuropathology, the degree of faulty signalling and connectivity is likely to increase, with further increases in variability. Taken together, this behavioural and neuroimaging research provides evidence that supports the proposal that IIV is a sensitive indicator of neurobiological integrity.

As well as indicating existing neuropathology, measures of IIV may also identify individuals who are at risk of future cognitive impairment, as increased IIV may be indicative of subthreshold impairment prior to broader cognitive decline. In addition, increased IIV may also predict mortality as there has long been evidence for accelerated cognitive deterioration in proximity to death (Riegel & Riegel, 1972). Given the weight of evidence, there is a pressing need to identify individuals who are at risk of adverse outcomes as this may allow interventions to target those most likely to benefit. Moreover, there is evidence that treatments for conditions such as Alzheimer's disease may be more beneficial if implemented early in the disease process. Therefore, in addition to considering changes in variability over time, this review also focuses on longitudinal studies that investigate the relationship between baseline IIV and important health-related outcomes such as cognitive impairment, dementia or mortality.

Do IIV Measures Possess Unique Properties?

Several methods have been used to compute IIV across RT trials, the most basic of which is the raw intraindividual *SD*. Other measures take into account systematic variance associated with influences such as time on task and experimental condition (e.g., adjusted intraindividual *SD*), or adjust for mean level of responding (e.g., the coefficient of variation: intraindividual *SD*/intraindividual mean RT). Occasionally, investigators have also used

metrics that reflect the typically non-normal distribution of RTs such as the interquartile range, or fitting the ex-Gaussian distribution. This produces three metrics, *mu*, *sigma* and *tau*, with the latter indexing intermittently slower responses that fall into the tail of the RT distribution.

There is typically a high correlation between measures of variability (e.g., intraindividual standard deviation) and central tendency (e.g., mean-RT) taken from the same cognitive task. It has been argued that increases in IIV associated with age (Myerson, Robertson, & Hale, 2007) and mild cognitive impairment (Phillips, Rogers, Haworth, Bayer, & Tales, 2013) reflect a general slowing of responses. By contrast, there is evidence that IIV provides greater differentiation in identifying persons with cognitive impairment (e.g., Dixon et al., 2007), and it has been shown that IIV is associated with brain imaging metrics such as white matter hyperintensities in early old age whereas mean-RT is not (e.g., Bunce et al., 2007). Therefore, a further consideration for this review is whether IIV measures predict outcome over and above associations accounted for by measures of central tendency.

Another important consideration is whether there is evidence that IIV adds to the predictive power of commonly used neuropsychological assessment measures. As such tasks are sensitive to both normal and neuropathological ageing, it is important to establish if IIV provides unique information that may help identification of persons at risk of adverse health outcomes. Finally, we explored whether findings varied systematically according to methodological differences across studies, such as task complexity, the number of RT trials, and how IIV was computed. As there is currently little consensus as to the optimum IIV measure, we considered if there was evidence favouring a particular type of IIV metric.

Methods

Studies were included in the present review if they appeared in an English language peer-reviewed journal, had a longitudinal design, used IIV measures at baseline, and the sample had a baseline mean age >55 years. This cut-off was selected in order to restrict the review to ageing research and also encompass early old-age. Studies were excluded if they were cross-sectional or the follow-up period was less than one year, as we were interested in longer-term change. Studies were also excluded if they did not have measures of RT IIV (i.e., they only reported accuracy or mean-RT, measured IIV across a battery of cognitive tasks, or investigated IIV in non-cognitive domains). Finally, as a systematic review of the association between IIV, falls and gait disturbance has recently been published by our group (Graveson, Bauermeister, McKeown, & Bunce, 2015), we excluded studies that had these outcomes.

Literature searches were completed on 17th May 2016 using the following electronic databases: EMBASE (from 1947, OVID interface), MEDLINE (from 1946, OVID interface), PsycINFO (from 1806, OVID interface), and Web of Science (from 1900). Three descriptors were used. First, IIV and common variants ('IIV', 'Intraindividual variability', 'intra-individual variability', 'within person variability', 'reaction time variability', 'response time variability', 'RT variability', 'reaction time inconsistency', 'response time inconsistency', 'RT inconsistency'), second, longitudinal research (longitudinal, 'cohort stud*', 'cohort design', 'follow up stud*', 'follow up design', 'prospective stud*', 'prospective research', 'prospective design', 'retrospective stud*', 'retrospective research', 'retrospective design'), and third, terms relating to ageing or potential longitudinal outcomes ('older adults', 'aging', 'ageing', 'neurocognitive disorder', 'neurodegenerative disease', 'dementia', 'Alzheimer's disease', 'mild cognitive impairment', 'Parkinson's disease', 'Huntington's disease', 'lewy body disorder', 'frontotemporal lobar degeneration', 'mortality', 'death'). All descriptors were searched for in keywords, title and abstract, and for the searches using EMBASE, MEDLINE, and PsycInfo, Medical Subject Heading (MeSH) terms were also used where

available. Additionally, citations in shortlisted papers and related review articles were inspected for further relevant studies.

Investigations were assessed according to the Quality in Prognostic Studies framework (Hayden, Cote, & Bombardier, 2006; Hayden, van der Windt, Cartwright, Cote, & Bombardier, 2013), which evaluates studies for bias and quality according to six criteria: Study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting. Consensus agreement was reached for each investigation. Where studies used overlapping samples or produced conflicting results, we drew our conclusions taking into account these criteria¹.

Results

The literature search initially identified 688 studies, of which 22 met the inclusion criteria (see Figure 1 for a flow diagram of the study selection process). After exclusions, there were nine studies that investigated change in variability over time and 17 studies that assessed whether IIV was associated with subsequent outcome (see Table 1 for details). Identified outcomes were cognitive change, conversion to mild cognitive impairment or dementia, and mortality. Three studies had a sample mean age <55 years (Bielak, Cherbuin, Bunce, & Anstey, 2014; Deary & Der, 2005b; Shipley, Der, Taylor, & Deary, 2006), but as these all reported analyses stratified by age group, data from the older group was included in this review. Although we identified 22 studies that met our inclusion criteria, several of these involved samples drawn from the same study population. Specifically, the PATH Through Life Study (Bielak et al., 2014; Das et al., 2014), Project MIND (Bielak, Hultsch, Strauss, MacDonald, & Hunter, 2010a, 2010b; Grand, Stawski, & MacDonald, 2016; Yao, Stawski, Hultsch, & MacDonald, 2016), the Victoria Longitudinal Study (Bielak, Hughes, Small, &

¹ Details of this evaluation are available from the corresponding author.

Dixon, 2007; MacDonald, Hultsch, & Dixon, 2003, 2008; Whitehead, Dixon, Hultsch, & MacDonald, 2011), the West of Scotland Twenty-07 Study (Deary & Der, 2005a, 2005b), while two further studies used identical participants (Bayer et al., 2014; Tales et al., 2012). The majority of these investigations fall into different sections below. However, where there was overlap, as noted earlier, we used formalized evaluation criteria in considering the findings.

In the following sections, first, we review studies of longitudinal change in IIV as a function of age, and then we turn to studies in which IIV serves as a predictor of future outcome. We then considered whether IIV was predictive of outcome over and above mean RT derived from the same task, and assess IIV relative to traditional neuropsychological tasks. Finally, we considered whether methodological differences across studies influenced the findings.

Longitudinal change in variability

With the exception of a study in Parkinson's disease patients (de Frias, Dixon, & Camicioli, 2012), all investigations drew participants from large-scale population-based studies of ageing. The majority of studies found that IIV increased over time (Bielak et al., 2014; Deary & Der, 2005b; Lovden, Li, Shing, & Lindenberger, 2007; MacDonald et al., 2003). This was consistently shown for old-old individuals (Lovden et al., 2007; MacDonald et al., 2003) and there was evidence that the rate of change increased above 75 years of age (Bielak et al., 2010b), although it was less clear at what age increases in IIV start. Two studies found significant eight-year increases in IIV for individuals who were aged 55 to 59 years (Deary & Der, 2005b) or 60 to 64 years (Bielak et al., 2014) at baseline. By contrast, Macdonald and colleagues (2003) only found six-year increase in IIV for participants aged over 75 years, with no increase evident for those aged 55-64 or 65-74 years. In another study, IIV did not increase over time in cognitively intact adults aged 65-84 years (de Frias, Dixon,

Fisher, & Camicioli, 2007). However, methodological differences may underlie this result as this study had a smaller sample, shorter follow-up duration (1.5 years), and only two assessment points compared with three or more in the other studies.

Additional factors influencing change in IIV that have been investigated included lifestyle activities (Bielak et al., 2007), Type II diabetes (Whitehead et al., 2011), and genetics (Das et al., 2014). However, results were not conclusive regarding the impact of these factors. For example, increasing IIV was associated with decreases in passive lifestyle activities such as reading newspapers and watching television, and novel activities including playing Bridge or completing a tax return, but not with physical or social activity, travel and self-maintenance (Bielak et al., 2007). Also, this result was found only for the most complex RT task, suggesting that the benefits of an engaged lifestyle are more likely to appear in more cognitively demanding tasks. Conversely, however, the influence of Type II diabetes on increasing variability over time was only evident in less demanding RT tasks (Whitehead et al., 2011). Lastly, although baseline IIV was influenced by genetic variants of Catechol-O-Methyltransferase (*COMT*) and Brain-Derived Neurotrophic Factor (*BDNF*), neither genotype primary effects nor interactions involving genotype affected the trajectory of change in IIV (Das et al., 2014).

Associations between variability and outcome

A total of 17 studies were identified that assessed relations between variability and outcomes that included cognitive change, mild cognitive impairment or dementia, and mortality. Of these, four studies also reported results for change in IIV over time, and were described in the preceding section (Bielak et al., 2010b; de Frias et al., 2012; Lovden et al., 2007; MacDonald et al., 2003).

Cognitive change. Six studies considered whether baseline IIV was related to change in cognitive performance in community-dwelling adults (Bielak et al., 2010b; Ghisletta, Fagot, Lecerf, & De Ribaupierre, 2013; Grand et al., 2016; Lovden et al., 2007; MacDonald et al., 2003; Yao et al., 2016). In an early study, adjusting for initial IIV accounted for cognitive change in several tasks including processing speed, memory and language (MacDonald et al., 2003). Similarly, Lovden and colleagues (2007) found that higher trial-to-trial variability both predicted and temporally preceded cognitive decline in verbal fluency and perceptual speed, whereas conversely, initial cognitive performance had a negligible influence on decline in IIV. Three additional studies investigated IIV and cognitive change using the same dataset, with a follow-up duration of three (Bielak et al., 2010b) or six years (Grand et al., 2016; Yao et al., 2016). In the first study, baseline IIV predicted the rate of cognitive change on tasks involving processing speed, fluency, reasoning, memory and language (Bielak et al., 2010b). More recent analyses over six years confirmed that baseline IIV moderated the rate of cognitive change, but only for memory and vocabulary and not processing speed, reasoning or fluency tasks (Grand et al., 2016). In addition, the rate of cognitive change increased with proximity to participant attrition, and higher IIV was associated with greater decline per year closer to attrition for memory and executive function (Trailmaking B: Yao et al., 2016). A final study investigated the amplitude of fluctuations in IIV and temporal dependency of responses (the relationship of each RT to the RTs that immediately preceded it). While the temporal dependency did not predict cognitive change (2-year follow-up adjusting for baseline scores), the amplitude of fluctuation in IIV was associated with change in fluid intelligence, but not with change in crystallised intelligence (Ghisletta et al., 2013). Taken together, these results suggest that baseline IIV predicts cognitive change across multiple cognitive domains, particularly for tasks assessing fluid abilities.

If fluctuations in RT are a marker for cognitive decline, there is a likelihood of covariation between change in IIV and change in cognitive performance. Consistent with this prediction, three of the aforementioned studies found significant covariation over time between IIV and cognition over one (Bielak et al., 2010b), two (Lovden et al., 2007) and three years (MacDonald et al., 2003). That is, IIV was inversely related to cognitive performance over time (higher IIV associated with lower cognitive performance), and greater baseline variability was related to more marked subsequent cognitive decline. In addition, when age was treated as a source of between-subject variance, it did not influence the covariation relationship (Bielak et al., 2010b; MacDonald et al., 2003) suggesting the association between IIV and cognitive ability is consistent across the older adult age range.

Mild Cognitive impairment and dementia. Seven studies investigated whether IIV predicted conversion to cognitive impairment or dementia over 1.5- to 10-year periods. All found that baseline IIV for at least one measure, was associated with outcome. In two large-scale population-based studies, higher initial IIV was associated with an increased risk of mild cognitive impairment over four- (Cherbuin, Sachdev, & Anstey, 2010) and five-year (Bielak et al., 2010a) intervals. However, in the latter study, IIV measures were not able to distinguish between persons who showed stable cognitive decline compared with those exhibiting fluctuating performance over time. It is unclear whether these groups had the same long-term prognosis, or whether those who show stable decline were further along the continuum of impairment, and were more likely to progress to dementia. Two studies looked at incident dementia in population-based samples. Baseline IIV (Kochan et al., 2016) and the ex-Gaussian parameter *tau* (Balota et al., 2010), both predicted dementia conversion over 4 or 10 years, respectively. Additionally, the association was weaker but remained significant after controlling for several established dementia risk factors. Further to these population-based studies, variability has also been assessed in relation to future dementia in clinical

samples. One such study assessed 1.5-year IIV change in Parkinson's disease patients who converted to dementia relative to non-converters and disease free controls (de Frias et al., 2012). Increased IIV over time was evident for dementia converters, but only distinguished the incipient Parkinson's dementia group from the cognitively intact group and not the Parkinson's disease dementia converters from non-converters. From a clinical perspective, the differentiation between Parkinson's dementia converters and non-converters is potentially the more important comparison if the aim is to detect individuals with Parkinson's disease who are more likely to cognitively deteriorate. Finally, baseline IIV was greater in patients with amnesic MCI who subsequently converted to dementia compared to non-converters and cognitively healthy older adults (Bayer et al., 2014; Tales et al., 2012). However, as converters were also more cognitively impaired at baseline, it is unclear whether these results would remain if absolute baseline differences in cognition were controlled for.

Mortality. Four studies examined whether IIV was associated with all-cause mortality. Three reported data from older-adult samples (Batterham, Bunce, Mackinnon, & Christensen, 2014; MacDonald et al., 2008; Shipley et al., 2006) and one from middle-aged participants (Deary & Der, 2005a). All of these studies found that greater baseline IIV was associated with an increased risk of death over periods of between 12 and 19 years. These associations remained statistically significant when controlling for influences typically associated with mortality such as socio-demographic factors, health behaviours and health status. In addition, one study investigated the terminal decline hypothesis by modelling the trajectories of longitudinal IIV change as a function of time to death. Decedents exhibited increased IIV for each year closer to death and this effect was larger for those aged 80 to 95 years relative to those aged 50 to 79 years (MacDonald et al., 2003).

Comparisons with other measures

In this section we consider the potential of IIV relative to either measures derived from the same RT task (e.g., mean or median RT). As IIV typically increases with response slowing, it is important to consider whether IIV effects are independent of more general slowing. Additionally, as variability measures may supplement current assessment methods, we assess whether IIV offers potential over and above traditional neuropsychological tasks.

Measures of central tendency. The majority of studies (13 out of 17) included measures of central tendency (mean or median RT; MRT) derived from the same cognitive task. Seven of these reported results from models that contained both IIV and MRT measures. Here, IIV was predictive of outcome over and above MRT in four studies (Batterham et al., 2014; Cherbuin et al., 2010; Ghisletta et al., 2013; MacDonald et al., 2008), although this was not universal, as two studies found superior performance for MRT (Kochan et al., 2016; Shipley et al., 2006). In the remaining study, there was comparable performance (Deary & Der, 2005a).

Other neuropsychological tasks. Six (out of 17) studies assessed the association between RT measures (IIV alone, or IIV and MRT added in a single step) and outcome relative to widely used neuropsychological tasks such as verbal recall or fluency, and IQ. In all studies, the results indicated that RT measures offered additional predictive utility over traditional tasks in predicting outcome, but there were mixed findings as to which RT measure was the best predictor. For example, relative to MRT and standardized cognitive tasks, IIV uniquely predicted MCI (Cherbuin et al., 2010). Additionally, a composite measure of IIV, verbal recall and sustained attention performed better than the widely used Mini-Mental State Examination (MMSE) in detecting MCI. Similarly, another investigation found that IIV produced a reliable increment in discrimination between dementia converters and non-converters in 14 out of 15 standardized psychomotor tasks (Balota et al., 2010). With regard to the association with mortality, greater IIV was associated with an increased risk of

death independent of demographic, cardiovascular disease, MRT and cognitive level measures (MacDonald et al., 2008). Two other studies used backwards elimination to find the best predictors of mortality. Here, simple IIV and complex MRT (Deary & Der, 2005a), and complex MRT (Shipley et al., 2006), were retained in the model, whereas IQ and memory or visuospatial reasoning were not. Finally, using receiver operating characteristic analyses, one further study found that combined RT measures (IIV and MRT) compared favourably with traditional neuropsychological measures in the prediction of incident dementia (Kochan et al., 2016).

Methodological differences across studies

There were methodological differences between studies in terms of the cognitive task or domain used to generate RTs, how many RT trials were included in the computation of metrics, and how IIV was calculated from the raw RTs. This raises the question of whether there is systematic variance in findings related to these factors, and whether there is evidence of an optimum IIV measure.

Task complexity. A dimension upon which tasks across studies varied, was cognitive complexity. Tasks ranged from those involving low cognitive demands (e.g., finger tapping) to more cognitively complex tasks (e.g., task switching). Thirteen (of 17) studies used multiple tasks or conditions of differing complexity, although two (Balota et al., 2010; MacDonald et al., 2003) reported results based on a single composite. The findings across studies suggested that tasks of varying complexity were all associated with outcomes including cognitive change (Grand et al., 2016; Yao et al., 2016), cognitive impairment (Bielak et al., 2010a; Cherbuin et al., 2010) and mortality (Batterham et al., 2014; Deary & Der, 2005a; MacDonald et al., 2008; Shipley et al., 2006). There were two exceptions, both of which only found significant effects for the more complex tasks (Bielak et al., 2010b; Kochan et al., 2016). However, this was relative to the other tasks within the study rather than

on the overall continuum of cognitive tasks. For example, in Kochan and colleagues' (2016) study, although choice-RT IIV and not simple-RT IIV predicted future dementia, the differences in complexity between the two tasks was relatively small, and both were less cognitively demanding than tasks used elsewhere. In addition, other investigations using the same cohort and measures as Bielak and colleagues (2010b), where stronger effects for more complex tasks were evident, found significant associations for both basic and complex composite measures (Bielak et al., 2010a; Yao et al., 2016). In sum, it appears that the findings for task complexity are inconsistent, and comparing across studies, the evidence suggests that moderately complex tasks are sufficient to generate reliable metrics of IIV.

Number of trials. The majority of studies used between 20 and 60 RT trials in computations of IIV, although one group repeat-tested the RT tasks over multiple weeks and averaged across sessions (Bielak et al., 2010a, 2010b; Yao et al., 2016). Across studies, there does not appear to be reliable differences arising from the number of RT trials used, or whether the testing session was repeated. Although there was insufficient evidence to determine whether there is an optimum number of trials, the studies included here suggest that 20 to 60 trials were sufficient to generate reliable IIV metrics that were associated with future outcome.

Measure of IIV. Studies also differed in how IIV metrics were computed. The majority used either the raw *SD*, or adjusted *SD* partialling out systematic variance associated with influences such as trial number, block, experimental condition and age group. Some investigations however, used measures that controlled for mean level of responding (e.g., coefficient of variation), fitted alternative models to individual RTs (e.g., an ex-Gaussian distribution), or used non-parametric measures (e.g., interquartile range). Where studies reported results from multiple metrics, these tended to show converging results, with IIV from all measures showing an association with outcome (e.g., Batterham et al., 2014;

Cherbuin et al., 2010; Lovden et al., 2007). Across studies, the different IIV metrics tended to produce similar results and there did not appear to be systematic variation according to the method used to compute the IIV metric.

Discussion

This is the first systematic review to consider longitudinal change in IIV and whether such measures predict cognitive and adverse health-related outcomes. The first main finding was that IIV increased over time, particularly in participants aged over 75 years. Second, greater baseline IIV was consistently associated with an increased risk of cognitive decline, cognitive impairment, dementia, or mortality. There was also evidence that this was independent of general slowing, and that IIV added to the predictive utility of commonly used neuropsychological assessment tools. Lastly, findings did not vary systematically according to methodological differences such as task complexity, number of trials and how IIV was calculated.

The review provided clear evidence that IIV increased over time in normal ageing. This is in line with cross-sectional studies and reviews that show higher IIV in older compared to younger groups (Dykiert et al., 2012), and in old-old relative to young-old individuals (Hultsch et al., 2002). There were, however, conflicting results as to the age at which increases in IIV began. While there were consistent findings of increasing IIV in old-old individuals, results varied across studies of younger individuals. Although theoretically it is plausible that age-related increases in variability begin in middle age (and perhaps earlier), further evidence supporting this possibility is clearly needed. Although beyond the scope of the present review, a meta-analysis of longitudinal change in IIV would help elucidate these inconsistent findings. In addition, three studies considered other influences on the trajectory of IIV change over time including both protective (lifestyle activities) and risk factors (Type II diabetes and *COMT* and *BDNF* gene variants) for broader cognitive decline. However,

none of these factors consistently influenced change in IIV. Together, the findings suggest that further work is needed to understand influences on age-related increases in IIV, and whether such increases are changeable, for example, through lifestyle intervention.

The review also found clear evidence that greater baseline IIV was associated with greater cognitive decline and an increased risk of MCI, dementia, or mortality over the follow-up period. The results are consistent with cross-sectional findings of increased IIV in populations with age-related conditions such as MCI and dementia (e.g., Christensen et al., 2005; Dixon et al., 2007; Gorus et al., 2008). Importantly however, as the majority of studies measured IIV at a time when participants were non-demented and cognitively intact, the results suggest that increases in IIV may precede these adverse health-related outcomes by several years. These findings highlights the potential for IIV as a prognostic measure that may help identify individuals at risk of future deleterious outcomes. They are also consistent with the proposal that IIV is a behavioural marker of neurobiological integrity (Hultsch et al., 2000; Hultsch et al., 2008). Variability may be sensitive to early neuropathological changes that influence attentional and executive control mechanisms (cf. Bunce et al., 2004; Bunce et al., 1993; West et al., 2002), which precede broader cognitive dysfunction. What is also not currently clear is whether increased IIV is related to specific neurological outcomes such as Alzheimer's disease, or is a universal sign of broader changes to brain integrity. Although it is important that future research address this question, it is of note that IIV also predicted all-cause mortality, suggesting that increased IIV is not necessarily specific to Alzheimer's pathology. IIV may predict mortality as neurobiological compromise may reflect broader biological processes that themselves portend to death. However, as IIV was also associated with an increased hazard of mortality in a middle-aged sample (Deary et al., 2006), the mechanism linking IIV and mortality may be complex and needs to be further elucidated.

An important consideration was whether IIV predicted outcome independently of estimates of central tendency. Measures such as standard deviation are typically highly correlated with mean RT taken from the same task and it has been argued that age-related differences in IIV reflect a general slowing of responses (e.g., Myerson et al., 2007). It is therefore appropriate to exercise caution in interpreting findings from studies that do not control for MRT. Nonetheless, across the reviewed studies, although greater baseline MRT was associated with an increased risk of cognitive impairment and mortality, IIV showed consistent results across studies, and tended to be associated with outcome after adjusting for MRT. The results suggest, therefore, that associations between IIV and outcome were not simply related to general slowing and that IIV measures possess unique predictive utility. This finding is consistent with cross-sectional studies that show differentiation between IIV and MRT (e.g., Bunce et al., 2007; Dixon et al., 2007).

Although assessed in fewer studies, the results also indicated that IIV predicted outcome over and above widely used neuropsychological tests. Again, this suggests that variability measures offer unique information relative to other commonly used tasks, and may therefore serve as a useful supplement to standard neuropsychological test batteries in healthcare settings. This is an area that would benefit from further research to determine the clinical utility of IIV, either alone or in combination with traditional measures. As an example, Cherbuin et al. (2010) showed that variability measures in combination with other cognitive tasks exhibited greater predictive power for cognitive impairment than the MMSE.

There were methodological variations across studies that included differences in RT tasks, the number of RT trials used, and how the IIV metric was computed. Previous research suggests there are greater differences associated with age (Dykiert et al., 2012) and dementia (de Frias et al., 2007; Gorus et al., 2008) using IIV measures from more demanding tasks. Additionally, a recent systematic review from our group suggested more sophisticated

measures of IIV may possess greater utility in understanding the relationship between IIV and falls or gait disturbance (Graveson et al., 2015). Nonetheless, in the present review we did not find any systematic differences in results according to task complexity, number of RT trials, or the IIV metric used. However, the present results indicate that moderately complex tasks, with 20 to 60 trials are sufficient to produce reliable IIV metrics. This is consistent with a recent cross-sectional study that found that 20 trials taking approximately 52 s to administer provided a reliable estimate of frontal white matter integrity (Bunce et al., 2013). The present findings also suggest that it is not necessary to use multiple testing sessions, or to compute mathematically complex measures of IIV. This is of note as such complexities would create practical difficulties when using the measures for neuropsychological assessment in clinical settings. Accordingly, the coefficient of variation may be appropriate in such settings, as it is relatively straightforward to compute and takes mean RT into account.

Although there were consistent findings across studies indicating that IIV measures were associated with future cognitive impairment or mortality, there are some caveats to this conclusion. First, while we identified 22 studies for inclusion in this review, several of these involved samples drawn from the same study population. This interdependence between some studies should be kept in mind when considering the findings of this review. Second, the samples were restricted to Western Europe, Australia, and North America and there is clearly a need for similar work in populations from Central and South America, Africa and Asia. Third, few studies reported metrics such as the sensitivity or specificity of IIV measures, though where these were reported (e.g., Cherbuin et al., 2010; de Frias et al., 2007), variability measures performed well. Lastly, none of the studies provided concrete cut-off scores for normative performance. The absence of normative standards limits the current clinical utility of IIV metrics. It is therefore of pressing importance that future research

determines normal ranges for IIV and IIV change in healthy older populations as this is central to developing the clinical utility of the measure.

Despite these considerations, having used a recognized framework to evaluate the reviewed studies (Hayden et al., 2006; Hayden et al., 2013), there are a number of strengths of the research that should be highlighted. First, the majority of investigations were population based studies with large sample sizes including a broad spectrum of individuals at risk of measured outcomes and clearly defined inclusion and exclusion criteria. Additionally, most studies had follow-up data on more than 80% of the sample and many used statistical procedures that took missing data into account. All of the studies included clear descriptions of how IIV was measured and how the outcome was defined. Lastly, many of the studies also controlled for potential confounding factors, including known risk factors for dementia and mortality.

In conclusion, the present review considered nine studies that looked at change in IIV over time and 17 studies that investigated whether IIV was associated with cognitive and adverse health-related outcomes. The results suggest that increasing IIV over time is related to normal ageing, but that more marked increases in IIV may indicate future cognitive decline, mild cognitive impairment, dementia, or mortality. One intriguing possibility is that IIV may provide a behavioural index of “brain frailty” arising from structural and physiological changes in the ageing brain. As such, measures of IIV may have considerable potential in clinical settings as they may supplement existing neuropsychological test batteries and help identify persons for early therapeutic intervention. Measures of IIV may therefore have considerable potential in clinical settings and offer practical advantages as administration is quick and requires little neuropsychological training. Additionally, the tasks often involve minimal linguistic content and may therefore be suitable for use with individuals from diverse backgrounds. Future work is needed to develop standardized

methods for measuring variability, with clear guidelines and norms for interpretation of results in community-based and clinical populations.

Acknowledgements

There are no conflicts of interest to declare.

References

- Balota, D. A., Tse, C. S., Hutchison, K. A., Spieler, D. H., Duchek, J. M., & Morris, J. C. (2010). Predicting conversion to dementia of the Alzheimer's type in a healthy control sample: the power of errors in Stroop color naming. *Psychology and Aging*, 25(1), 208-218. doi: 10.1037/a0017474
- Batterham, P. J., Bunce, D., Mackinnon, A. J., & Christensen, H. (2014). Intra-individual reaction time variability and all-cause mortality over 17 years: a community-based cohort study. *Age and Ageing*, 43(1), 84-90. doi: 10.1093/ageing/aft116
- Bayer, A., Phillips, M., Porter, G., Leonards, U., Bompas, A., & Tales, A. (2014). Abnormal inhibition of return in mild cognitive impairment: is it specific to the presence of prodromal dementia? *Journal of Alzheimers Disease*, 40(1), 177-189. doi: 10.3233/JAD-131934
- Bielak, A. A., Cherbuin, N., Bunce, D., & Anstey, K. J. (2014). Intraindividual variability is a fundamental phenomenon of aging: evidence from an 8-year longitudinal study across young, middle, and older adulthood. *Developmental Psychology*, 50(1), 143-151. doi: 10.1037/a0032650
- Bielak, A. A., 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 Series B-Psychological Sciences and Social Sciences*, 62(6), P331-P339.
- Bielak, A. A., Hultsch, D. F., Strauss, E., MacDonald, S. W. S., & Hunter, M. A. (2010a). Intraindividual Variability in Reaction Time Predicts Cognitive Outcomes 5 Years Later. *Neuropsychology*, 24(6), 731-741. doi: 10.1037/a0019802

Bielak, A. A., Hultsch, D. F., Strauss, E., MacDonald, S. W. S., & Hunter, M. A. (2010b).

Intraindividual Variability Is Related to Cognitive Change in Older Adults: Evidence for Within-Person Coupling. *Psychology and Aging*, 25(3), 575-586. doi:

10.1037/a0019503

Bunce, D., Anstey, K. J., Cherbuin, N., Burns, R., Christensen, H., Wen, W., & Sachdev, P.

S. (2010). Cognitive Deficits Are Associated with Frontal and Temporal Lobe White Matter Lesions in Middle-Aged Adults Living in the Community. *Plos One*, 5(10).

doi: ARTN e13567

10.1371/journal.pone.0013567

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(9), 2009-2015. doi:

10.1016/j.neuropsychologia.2007.02.006

Bunce, D., Bielak, A. A. M., Cherbuin, N., Batterham, P. J., Wen, W., Sachdev, P., &

Anstey, K. J. (2013). Utility of Intraindividual Reaction Time Variability to Predict White Matter Hyperintensities: A Potential Assessment Tool for Clinical Contexts?

Journal of the International Neuropsychological Society, 19(9), 971-976. doi: Doi

10.1017/S1355617713000830

Bunce, D., MacDonald, S. W. S., & Hultsch, D. F. (2004). Inconsistency in serial choice

decision and motor reaction times dissociate in younger and older adults. *Brain and Cognition*, 56(3), 320-327. doi: DOI 10.1016/j.bandc.2004.08.006

Bunce, D., Warr, P. B., & Cochrane, T. (1993). Blocks in Choice Responding as a Function

of Age and Physical-Fitness. *Psychology and Aging*, 8(1), 26-33. doi: Doi

10.1037/0882-7974.8.1.26

- Cherbuin, N., Sachdev, P., & Anstey, K. J. (2010). Neuropsychological Predictors of Transition From Healthy Cognitive Aging to Mild Cognitive Impairment: The PATH Through Life Study. *American Journal of Geriatric Psychiatry*, 18(8), 723-733. doi: 10.1097/Jgp.0b013e3181cdecf1
- 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(3), 309-317. doi: 10.1037/0894-4105.19.3.309
- Das, D., Tan, X., Bielak, A. A., Cherbuin, N., Easteal, S., & Anstey, K. J. (2014). Cognitive ability, intraindividual variability, and common genetic variants of catechol-O-methyltransferase and brain-derived neurotrophic factor: a longitudinal study in a population-based sample of older adults. *Psychology and Aging*, 29(2), 393-403. doi: 10.1037/a0035702
- de Frias, C. M., Dixon, R. A., & Camicioli, R. (2012). Neurocognitive Speed and Inconsistency in Parkinson's Disease with and without Incipient Dementia: An 18-Month Prospective Cohort Study. *Journal of the International Neuropsychological Society*, 18(4), 764-772. doi: 10.1017/S1355617712000422
- de Frias, C. M., Dixon, R. A., Fisher, N., & Camicioli, R. (2007). Intraindividual variability in neurocognitive speed: A comparison of Parkinson's disease and normal older adults. *Neuropsychologia*, 45(11), 2499-2507. doi: 10.1016/j.neuropsychologia.2007.03.022
- Deary, I. J., Bastin, M. E., Pattie, A., Clayden, J. D., Whalley, L. J., Starr, J. M., & Wardlaw, J. M. (2006). White matter integrity and cognition in childhood and old age. *Neurology*, 66(4), 505-512. doi: 10.1212/01.wnl.0000199954.81900.e2

- Deary, I. J., & Der, G. (2005a). Reaction time explains IQ's association with death. *Psychological Science*, 16(1), 64-69. doi: 10.1111/j.0956-7976.2005.00781.x
- Deary, I. J., & Der, G. (2005b). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging Neuropsychology and Cognition*, 12(2), 187-215. doi: 10.1080/138255805990969235
- Dixon, R. A., Lentz, T. L., Garrett, D. D., MacDonald, S. W. S., Strauss, E., & Hultsch, D. F. (2007). Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. *Neuropsychology*, 21(3), 381-399. doi: 10.1037/0894-4105.21.3.381
- Duchek, J. M., Balota, D. A., Tse, C. S., Holtzman, D. M., Fagan, A. M., & Goate, A. M. (2009). The Utility of Intraindividual Variability in Selective Attention Tasks as an Early Marker for Alzheimer's Disease. *Neuropsychology*, 23(6), 746-758. doi: 10.1037/a0016583
- Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Age differences in intra-individual variability in simple and choice reaction time: systematic review and meta-analysis. *Plos One*, 7(10), e45759. doi: 10.1371/journal.pone.0045759
- Fjell, A. M., Westlye, L. T., Amlien, I. K., & Walhovd, K. B. (2011). Reduced White Matter Integrity Is Related to Cognitive Instability. *Journal of Neuroscience*, 31(49), 18060-18072. doi: 10.1523/Jneurosci.4735-11.2011
- Ghisletta, P., Fagot, D., Lecerf, T., & De Ribaupierre, A. (2013). Amplitude of fluctuations and temporal dependency in intraindividual variability. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry*, 26(3), 141-151.
- Gorus, E., De Raedt, R., Lambert, M., Lemper, J. C., & Mets, T. (2008). Reaction times and performance variability in normal aging, mild cognitive impairment, and Alzheimer's

- disease. *J Geriatr Psychiatry Neurol*, 21(3), 204-218. doi: 10.1177/0891988708320973
- Grand, J. H., Stawski, R. S., & MacDonald, S. W. (2016). Comparing individual differences in inconsistency and plasticity as predictors of cognitive function in older adults. *J Clin Exp Neuropsychol*, 38(5), 534-550. doi: 10.1080/13803395.2015.1136598
- Graveson, J., Bauermeister, S., McKeown, D., & Bunce, D. (2015). Intraindividual reaction time variability, falls and gait in old age: A systematic review. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*. doi: 10.1093/geronb/gbv027
- Hayden, J. A., Cote, P., & Bombardier, C. (2006). Evaluation of the quality of prognosis studies in systematic reviews. *Annals of Internal Medicine*, 144(6), 427-437.
- Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Cote, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Ann Intern Med*, 158(4), 280-286. doi: 10.7326/0003-4819-158-4-201302190-00009
- Hultsch, D. F., MacDonald, S. W., 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(4), 588-598.
- Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, 57(2), 101-115.
- Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. S. (2008). Intraindividual variability, cognition and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The*

handbook of aging and cognition (3rd ed., pp. 491-556). New York: Psychology Press.

Jackson, J. D., Balota, D. A., Duchek, J. M., & Head, D. (2012). White matter integrity and reaction time intraindividual variability in healthy aging and early-stage Alzheimer disease. *Neuropsychologia*, 50(3), 357-366. doi: 10.1016/j.neuropsychologia.2011.11.024

Kochan, N. A., Bunce, D., Pont, S., Crawford, J. D., Brodaty, H., & Sachdev, P. S. (2016). Reaction Time Measures Predict Incident Dementia in Community-Living Older Adults: The Sydney Memory and Ageing Study. *Am J Geriatr Psychiatry*, 24(3), 221-231. doi: 10.1016/j.jagp.2015.12.005

Li, S. C., Lindenberger, U., & Sikstrom, S. (2001). Aging cognition: from neuromodulation to representation. *Trends in Cognitive Sciences*, 5(11), 479-486. doi: 10.1016/S1364-6613(00)01769-1

Lovden, 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(12), 2827-2838. doi: 10.1016/j.neuropsychologia.2007.05.005

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(3), 510-523. doi: 10.1037/0882-7974.18.3.510

MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2008). Predicting impending death: Inconsistency in speed is a selective and early marker. *Psychology and Aging*, 23(3), 595-607. doi: 10.1037/0882-7974.23.3.595

MacDonald, S. W. S., Karlsson, S., Rieckmann, A., Nyberg, L., & Backman, L. (2012).

Aging-Related Increases in Behavioral Variability: Relations to Losses of Dopamine

D-1 Receptors. *Journal of Neuroscience*, 32(24), 8186-8191. doi: Doi

10.1523/Jneurosci.5474-11.2012

Mella, N., de Ribaupierre, S., Eagleson, R., & de Ribaupierre, A. (2013). Cognitive

Intraindividual Variability and White Matter Integrity in Aging. *Scientific World*

Journal. doi: 10.1155/2013/350623

Moy, G., Millet, P., Haller, S., Baudois, S., de Bilbao, F., Weber, K., . . . Delaloye, C. (2011).

Magnetic resonance imaging determinants of intraindividual variability in the elderly:

combined analysis of grey and white matter. *Neuroscience*, 186, 88-93. doi:

10.1016/j.neuroscience.2011.04.028

Myerson, J., Robertson, S., & Hale, S. (2007). Aging and intraindividual variability in

performance: Analyses of response time distributions. *Journal of the Experimental*

Analysis of Behavior, 88(3), 319-337. doi: 10.1901/jeab.2007.88-319

Phillips, M., Rogers, P., Haworth, J., Bayer, A., & Tales, A. (2013). Intra-individual reaction

time variability in mild cognitive impairment and Alzheimer's disease: gender,

processing load and speed factors. *Plos One*, 8(6), e65712. doi:

10.1371/journal.pone.0065712

Riegel, K. F., & Riegel, R. M. (1972). Development, Drop, and Death. *Developmental*

Psychology, 6(2), 306-319. doi: 10.1037/H0032104

Shipley, B. A., Der, G., Taylor, M. D., & Deary, I. J. (2006). Cognition and all-cause

mortality across the entire adult age range: Health and lifestyle survey. *Psychosomatic*

Medicine, 68(1), 17-24. doi: 10.1097/01.psy.0000195867.66643.0f

- Tales, A., Leonards, U., Bompas, A., Snowden, R. J., Philips, M., Porter, G., . . . Bayer, A. (2012). Intra-Individual Reaction Time Variability in Amnesic Mild Cognitive Impairment: A Precursor to Dementia? *Journal of Alzheimers Disease*, 32(2), 457-466. doi: 10.3233/Jad-2012-120505
- Walhovd, K. B., & Fjell, A. M. (2007). White matter volume predicts reaction time instability. *Neuropsychologia*, 45(10), 2277-2284. doi: 10.1016/j.neuropsychologia.2007.02.022
- 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(3), 402-419. doi: 10.1006/brcg.2001.1507
- Whitehead, B. P., Dixon, R. A., Hultsch, D. F., & MacDonald, S. W. (2011). Are neurocognitive speed and inconsistency similarly affected in type 2 diabetes? *J Clin Exp Neuropsychol*, 33(6), 647-657. doi: 10.1080/13803395.2010.547845
- 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(1), 88-96. doi: 10.1037/0894-4105.19.1.88
- Yao, C., Stawski, R. S., Hultsch, D. F., & MacDonald, S. W. S. (2016). Selective attrition and intraindividual variability in response time moderate cognitive change. *J Clin Exp Neuropsychol*, 38(2), 227-237. doi: 10.1080/13803395.2015.1102869

Table 1. Summary of studies included in the review

| Reference | Population, sample size (%) ¹ , and age ² | Test measure(s) | Follow-up period | Outcome measure(s) | Main findings for IIV |
|------------------------|--|---|------------------|---|--|
| Change in IIV | | | | | |
| Bielak, et al., 2014 | Community dwelling older adults (n=2115, aged 60-64 years) | IIV – ISD for SRT and 2CRT | 8 years | Change in IIV | ISD increased over time (SRT, $\beta=.17$, $p<.001$; CRT, $\beta=.16$, $p<.001$) |
| Bielak, et al., 2007 | Community dwelling older adults (n=530, aged 55-94 years) | IIV –ISD for SRT, 4CRT, lexical decision, and semantic decision Other - Mean-RT Lifestyle activities – physical, self-maintenance, social, travel, passive, integrative, and novel information processing | 6 years | Change in IIV as a function of lifestyle activities | Change in lifestyle activities was associated with change in semantic decision ISD ($R^2=.03$, $p=.006$), with passive information processing ($\beta=-.15$, $p<.01$) and novel information processing ($\beta=-.10$, $p<.05$) as unique predictors. All other Models were non-significant. |
| Bielak et al., 2010b | Community dwelling older adults (n=304, aged 64-92) | IIV – ISD for Finger tapping (Motor composite), 4CRT, shape RT, colour RT (basic composite), CRT 1-back, task switching (complex composite) Other - mean-RT | 3 years | Change in IIV | In full sample, none of the ISD composites increased over time (average slopes did not differ from zero). For Motor ISD, there were no age differences on rate of change. For Basic ISD and Complex ISD, the rate of change increased with age over 75 years (slope increase of .03 for every year over 75, $p_{diff}<.001$) |
| Das et al., 2014 | Community dwelling older adults (n=400, aged 60-64 years) | IIV – ISD for SRT and 2CRT Other - mean-RT Genotype - <i>COMT*MET-</i> or <i>COMT*MET+</i> , <i>BDNF*MET-</i> or <i>BDNF*MET+</i> | 8 years | Change IIV as a function of genotype | ISD increased over time, but the trajectory of change was not different for the genotype groups. |
| Deary & Der, 2005b | Community dwelling older adults (n=673, aged 54-59 years) | IIV – raw SD for SRT and CRT; SD adjusted for mean-RT Other- Mean-RT | 8 years | Change in IIV | Using raw SD, IIV increased over time (statistics not reported) Using SD adjusted for mean-RT, IIV remained stable or declined slightly. |
| de Frias et al., 2007 | PD dementia converters (n=10, aged 66-84 years) non-converters (n=21, aged 65-77) and CH adults (n=43, aged 65-84 years) | IIV –ISD for SRT, 2CRT, 4CRT, 8CRT Other – mean-RT | 1.5 years | Change in IIV | Significant Group x Time interaction ($p<.01$); PD dementia converters became more variable over time while the non-converters and CD groups did not. |
| Lovden et al., 2007 | Community-dwelling older adults (n=447, aged 70-102 years) | IIV – CV for Identical Pictures Other - Mean-RT | 13 years | Change in IIV | There was an average 2-year increase in CV ($p<.05$). Older age reliably predicted greater change in variability. |
| MacDonald et al., 2003 | Community dwelling older adults (n=446 (59%), aged 55-89 years) | IIV – ISD for SRT, 2CRT, 4CRT, 8CRT, lexical decision, and semantic decision | 6 years | Change in IIV | ISD increased over time for semantic decision but declined for CRT. There were significant Age x Wave interactions for SRT ($p<.05$), lexical decision ($p<.01$), and semantic decision ($p<.01$). ISD remained |

| Reference | Population, sample size (%) ¹ , and age ² | Test measure(s) | Follow-up period | Outcome measure(s) | Main findings for IIV |
|-----------------------------------|---|---|------------------|--|---|
| | | | | | stable for those aged 54-65 and 64-75 years, and increased for individuals aged 75-89 years. |
| Whitehead et al., 2011 | Community-dwelling older adults (N=330 (71%), aged 53-90 years) with type II diabetes (n=30) and CH controls (n=302). | IIV – ISD for SRT, 2CRT, 4CRT, 8CRT, lexical decision, and semantic decision Other – mean-RT | 3 years | Change in IIV | There were significant effects of time on ISD for 2CRT ($p<.001$ (one tailed)), lexical decision ($p<.05$ (one tailed)), and semantic decision ($p<.05$ (one tailed)). Time x type II diabetes interactions were only evident for 2CRT ($p<.05$). |
| Outcome – Cognitive change | | | | | |
| Bielak et al., 2010b | Community dwelling older adults (n=304, aged 64-92) | IIV – ISD for Finger tapping (Motor composite), 4CRT, shape RT, colour RT (basic composite), 4CRT 1-back, task switching (complex composite) Other - mean-RT | 3 years | Change in cognitive functioning (digit symbol, letter series, word recall, similarities, vocabulary) | Complex ISD composite predicted cognitive change (digit symbol, $p<.001$; letter series, $p<.01$; word recall, $p<.01$, vocabulary, $p<.05$) There was significant 1-year covariance ($p<.01$) between ISD composites and cognitive measures in 10 of 15 models; When IIV was high, cognitive performance was correspondingly low. |
| Ghisletta et al., 2013 | Community dwelling older adults (n=165, aged 59-89) | IIV – detrended ISD ($ISDe_{i-1}$), autoregressive coefficient (ϕ_1) for SRT Other - Mean-RT | 2 years | Change in fluid and crystallized intelligence (Raven's progressive matrices and Mill Hill vocabulary) | $ISDe_{i-1}$ predicted Raven score at Time 2 when controlling for Raven score at Time 1 ($p=.01$). No significant association for ϕ_1 , or Mill Hill score at Time 2 controlling for Time 1 score. |
| Grand et al., 2016 | Community dwelling older adults (n=304, aged 64-92) | IIV – ISD for 4CRT, 4CRT 1-back Other – slope of change in ISD over five testing session two-weeks apart. | 6 years | Change in cognitive functioning (digit symbol, letter series, word recall, similarities, vocabulary) | Baseline 4CRT ISD moderated rates of cognitive change for word recall and vocabulary ($p<.05$). 4CRT 1-back ISD moderated rate of change for word recall only ($p<.05$). |
| Lovden et al., 2007 | Community-dwelling older adults (n=447, aged 70-102 years) | IIV – CV for Identical Pictures Other - Mean-RT | 13 years | Change in cognitive functioning (digit letters and verbal fluency) | Higher trial-to-trial variability preceded and predicted greater cognitive decline in perceptual speed and verbal fluency Significant correlations between 2-year change in IIV and change in digit letters ($r=-.68$, $p<.05$) and fluency ($r=-.82$, $p<.05$) |
| MacDonald et al., 2003 | Community dwelling older adults (n=446 (59%), aged 55-89 years) | IIV – ISD for SRT, 2CRT, 4CRT, 8CRT, lexical decision, and semantic decision | 6 years | Change in cognitive functioning (identical pictures, computation span, letter series, word recall, story recall, vocabulary) | Controlling for baseline IIV attenuated the variance associated with cognitive change (between 85.7 and 96.5% attenuation). Significant 3-year time-varying covariation was observed. Increasing inconsistency was associated with declining cognitive performance for 5 of 6 measures. |
| Yao et al., 2016 | Community dwelling older adults (n=304, aged 64-92) | IIV – ISD for 4CRT, shape RT, colour RT (basic composite), 4CRT 1-back, task switching (complex composite) | 6 years | Change in cognitive functioning (TMT parts A and B, free recall, WAIS III block design, vocabulary) | Higher IIV was associated with greater decline per year closer to study attrition (Basic ISD composite, word recall $p<.001$; Complex ISD composite, word recall $p<.001$, TMT-B $p<.001$). Other outcomes were non-significant. |

| Reference | Population, sample size (%) ¹ , and age ² | Test measure(s) | Follow-up period | Outcome measure(s) | Main findings for IIV |
|---|---|---|---------------------------|--|--|
| Outcome – Cognitive impairment or dementia | | | | | |
| Balota et al., 2010 | Healthy older adults (n=37 (100%) classified as converters (n=12, M=81.7 years) and non-converters (n=25, M=77.3 years) | IIV - Ex-Gaussian <i>tau</i> parameter for Stroop task (congruent and incongruent combined) Other – Mean-RT, Errors for Stroop task, standardized cognitive tests ³ | Average of 7.6 – 10 years | Dementia of Alzheimer's type (CDR≥0.5). | Using Cox Proportional Hazard Analysis <i>tau</i> was associated with dementia conversion, $p<.01$; OR 1.009, 95% CI=1.003-1.012). <i>Tau</i> produced a reliable increment in discrimination between converters and non-converters over 14/15 of the psychometric measures ($p<0.05$) |
| Bayer et al., 2014 | Patients with amnesic MCI (n=39 (87%) classified as converters (n=13, M=75±7 years) and non-converters (n=26, M=72±6 years) and cognitively healthy adults (n=31, M=73±5.0 years) | IIV-IQR for Posner exogenous cuing task Other – Median RT, attentional disengagement, inhibition of return, MMSE. | 2.5 years | Dementia (NINCDS-ADRDA criteria) | Raised baseline IQR in amnesic MCI converters compared with CH adults ($p<.026$) and non-converters ($p<.008$) for all cue target intervals. |
| Bielak et al., 2010a | Community-dwelling older adults (n=212 (70%) aged 64-92 years) | IIV – ISD for Finger tapping (Motor composite), 4CRT, shape RT, colour RT (basic composite), 4CRT 1-back, task switching (complex composite) Other - mean-RT | 5 years | Change in cognitive status (stable intact, fluctuating, stable decline and stable CIND-cognitive impairment no dementia) | Baseline ISD was raised in the fluctuating and declining groups. Basic ISD, fluctuating vs intact ($p<.05$; OR 1.29) declining vs intact ($p<.01$, OR=1.61). Complex ISD, fluctuating vs intact ($p<.01$; OR 1.20) declining vs intact ($p<.01$; OR=1.32). |
| Cherbuin et al., 2010 | Community-dwelling older adults (n=2082 (87%) aged 60-64 years) | IIV – Mean absolute residuals (MAR) and mean independent residuals (MIV) for SRT and CRT Other - mean-RT for SRT and CRT, standardized cognitive tasks ⁴ | 4 years | MCI; CDR≥0.5; any mild cognitive deficits (any-MCD) | SRT MAR ($p=.001$), SRT MIV ($p<.002$) and CRT MIV ($p<.046$) were all associated with MCI, CDR 0.5, and any-MCD. In the fully adjusted Model (all cognitive tasks, MRT and MAR) SRT MAR was associated with MCI ($p<.001$; OR 9.27, 95% CI=1.04-82.52), CDR 0.5 ($p=.008$; OR 10.85, 95% CI=1.82-64.39) and any-MCD ($p=.003$; OR 7.52, 95% CI=1.96-28.89). |
| de Frias et al., 2012 | PD patients (n=31 (100%) aged 65-84 years) classified as converters (n=10) or non-converters (n=21), and cognitively healthy adults (n=43, aged 65-84 years) | IIV – 18-month change in ISD for SRT, 2CRT, 4CRT, 8CRT Other – 18-month change in mean-RT | 1.5 years | Dementia or cognitive impairment | Significant Group x Time interaction ($p<.01$); PD dementia converters became more variable over time while the other groups did not. Using logistic regression, change in SRT ISD was a predictor of group membership between PD converters and CH ($p<.01$; OR 1.38, 95% CI=1.05-1.83) but not between PD converters and non-converters. |
| Kochan et al., 2016 | Community dwelling older adults (n=861 (83%), aged 70-90 years) | IIV – ISD for SRT and CRT Other – Mean-RT, Standardized cognitive tests ⁵ | 4 years | Dementia | Greater complex IIV was associated with an increased hazard of incident dementia ($p=.003$; HR=1.43). This remained significant when adjusting for clinical or medical covariates in separate models. After adjusting for mean-RT, incremental effects of IIV were not significant. |
| Tales et al., 2012 | Same sample at Bayer et al. (2014) | IIV - IQR for visual attention Other - Median RT and phasic alerting (no cue-cued RT) | 2.5 years | Dementia (NINCDS-ADRDA criteria) | Raised IQR in the amnesic MCI converters compared with non-converters ($p=.001$). There was no significant difference between aMCI non-converters and the CH group. |

| Reference | Population, sample size (%) ¹ , and age ² | Test measure(s) | Follow-up period | Outcome measure(s) | Main findings for IIV |
|----------------------------|--|---|------------------|---------------------|---|
| Outcome – Mortality | | | | | |
| Batterham et al, 2014 | Community based older adults (n=825 for SRT and n=802 for CRT, aged 70-97) | IIV – raw SD, CV, and ISD for SRT and 2CRT Other – mean-RT | 17 years | All-cause mortality | Greater IIV (all measures) was associated with an increased hazard of mortality ($p<.001$, $HR<1.14$). This remained significant for all SRT measures ($p\leq.01$) and for CRT raw SD and ISD ($p\leq.05$) when adjusting for covariates. The association remained significant for SRT IIV measures ($p\leq.02$) when additionally adjusting for mean-RT. |
| Deary & Der, 2005a | Community dwelling middle-aged adults (n=898, aged 54-59) | IIV – raw SD for SRT and 4CRT Other - mean-RT | 14 years | All-cause mortality | Greater raw SD was associated with an increased hazard of mortality (SRT, $p<.0001$, $HR=1.29$, 95% CI=1.15-1.44; CRT, $p=.023$, $HR=1.17$, 95% CI=1.02-1.34). Effects remained significant when adjusting for covariates (SRT, $p<.0001$, CRT, $p=.048$). There were no independent effects in fully adjusted model (mean-RT, raw SD, IQ and covariates). Using backwards elimination SRT SD and CRT mean-RT were retained in the model. |
| MacDonald et al., 2008 | Community dwelling older adults (n=707 (70%), aged 55-85 years) | IIV – ISD for lexical decision and semantic decision. Other- mean-RT, Cognitive performance level ⁶ | 12 years | All-cause mortality | Greater ISD in decedents compared with survivors ($p<.001$). ISD predicted mortality when adjusting for covariates and all cognitive tasks (Semantic decision, $p<.01$, lexical decision, $p<.05$). Semantic decision also predicted mortality after adjusting for mean-RT ($p\leq.05$) |
| Shipley et al., 2006 | Community dwelling adults (n=7,414 aged 18-99 years, n=963 aged 60+ years) | IIV – SD for SRT and 4CRT Other – mean-RT, verbal declarative memory and visuospatial reasoning | 19 years | All-cause Mortality | Controlling for age and sex, SRT ($HR=1.09$) and CRT ($HR=1.07$) raw SD was associated with mortality in 60+ group. This remained significant for SRT ($HR=1.03$) when adjusting for all covariates. Using backwards elimination (SD, mean-RT, memory, and reasoning) CRT mean-RRT was the only predictor that was retained. |

¹ Longitudinal sample size; (%) refers to percentage of baseline sample included in analyses. ² Age at baseline

³ Logical memory, digit span forward and backward, trail making test parts A and B, WAIS-R information, WIAS-R block design, WAIS-R digit symbol, Benton delay, Benton copy, Boston naming test, crossing off, mental control, associate recall, letter fluency

⁴ MMSE, immediate and delayed recall, digit span backwards, spot-the-word, symbol digit modality test

⁵ Logical memory - delayed, Rey auditory verbal learning test- delayed, category fluency, coding, WAIS-R block design, Benton visual retention, trailmaking test parts A and B, Boston naming test, letter fluency

⁶ Working memory (sentence construction, listening span, computation span), episodic memory (word list recall, story recall), semantic memory (world fact recall, recognition vocabulary)

BDNF= Brain-Derived Neurotropic Factor; *CDR*=clinical dementia rating; *CH*=cognitively healthy; 95% CI=95% confidence interval; *COMT*= Catechol-O-Methyltransferase; *CRT*=complex reaction time; *CV*=coefficient of variation; *HR*=hazard ratio; *IIV*=intraindividual variability; *ISD*=intraindividual standard deviation; *IQR*=interquartile range; *MCI*=Mild Cognitive Impairment; *MMSE*=Mini Mental State Examination; *PD*=Parkinson's disease; *SD*=standard deviation; *SRT*=simple reaction time; *TMT*=trailmaking test.

List of Figure legends

Figure 1: Flow diagram of the study selection process

Figure 1:

