RESEARCH REPOSITORY

This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination. The definitive version is available at:

https://www.apa.org/pubs/journals/neu

Mumme, R., Pushpanathan, M., Donaldson, S., Weinborn, M., Rainey-Smith, S.R., Maruff, P. and Bucks, R. Longitudinal association of intraindividual variability with cognitive decline and dementia: A meta-analysis. Neuropsychology

https://researchrepository.murdoch.edu.au/id/eprint/60666

Copyright: © 2021 American Psychological Association
It is posted here for your personal use. No further distribution is permitted.

Longitudinal association of intraindividual variability with cognitive decline and dementia: A meta-analysis.

Authors: Rachael Mumme, Maria Pushpanathan, Sara Donaldson, Michael Weinborn, Stephanie R. Rainey-Smith, Paul Maruff & Romola S. Bucks.

Conceptualisation: R.M, R.S.B, M.W, S.R.R-S, P.M. Data Curation: R.M, S.D, M.P. Formal Analysis: R.M, R.S.B. Investigation: R.M, S.D. Methodology: R.M, M.P, R.S.B, M.W. Project Administration: R.M. Resources: R.S.B, M.W, P.M. Software: R.B.S, M.P. Supervision: R.S.B, M.W, S.R.R-S Validation: R.S.B. Visualisation: R.M, R.S.B. Writing – original draft: R.M. Writing – review and editing: R.M, M.P, R.S.B, M.W, P.M, S.D, S.R.R-S.

Corresponding Author:

Romola S. Bucks, BSc. (Hons), MSc Clin Psych, PhD, FAPS
Director, The Raine Study
School of Psychological Science & School of Population and Global Health
University of Western Australia
35 Stirling Highway, Perth, WA 6009
+61 8 6488 3232
romola.bucks@uwa.edu.au

Authors Note: R.M was funded with the Australian Research Training Program (RTP) Stipend.

Abstract

Objective: Intraindividual variability (IIV) –variance in an individuals' cognitive performance - may be associated with subsequent cognitive decline and/or conversion to dementia in older adults. This novel measure of cognition encompasses two main operationalisations: inconsistency (IIV-I) and dispersion (IIV-D), referring to variance within or across tasks respectively. Each operationalisation can also be measured with or without covariates. This meta-analytic study explores the association between IIV and subsequent cognitive outcomes regardless of operational definitions and measurement approaches. **Method:** Longitudinal studies (N = 13) that have examined IIV in association with later cognitive decline and/or conversation to MCI/dementia were analysed. The effect of IIV operationalisation was explored. Additional sub group analysis of measurement approaches could not be examined due to the limited number of appropriate studies available for inclusion. **Results:** Meta-analytic estimates suggest IIV is associated with subsequent cognitive decline and/or conversion to MCI/dementia (r = .20 , 95% CI [.09, .31]) with no significant difference between the two operationalisations observed (Q = 3.41, p = .065). **Conclusion:** Cognitive IIV, including both IIV-I and IIV-D operationalisations, appears to be associated with subsequent cognitive decline and/or dementia and may offer a novel indicator of incipient dementia in both clinical and research settings.

Key Words: Intraindividual Variability, Cognitive decline, dementia.

Key Points

Question: Is Intraindividual Variability (IIV) and its alternative operationalisations associated with subsequent cognitive decline or conversion to dementia?

Findings: Greater IIV was associated with a higher risk of subsequent cognitive decline and/or conversation to MCI/dementia with no significant differences in this association seen across different approaches to measuring IIV.

Importance: These findings are useful in identifying a novel cognitive marker of subsequent cognitive decline and/or conversion to MCI/dementia in older adults.

Next Steps: The clinical utility of this measure should be further examined.

Longitudinal association of intraindividual variability with cognitive decline and dementia: A meta-analysis.

Much neuropsychological research seeks to understand brain behaviour relationships in dementia using total or average task scores on standardized tests of cognition. Over recent years, there has been increasing interest in the extent to which an individual's consistency of performance can also inform these models. Cognitive variability or intraindividual variability (IIV), referring to cognitive performance variability within a single individual, has been proposed as a potential early marker of dementia, including the dementia characteristic of Alzheimer's disease (AD) (Christ, Combrinck, & Thomas, 2018b; Bayer, & Tales, 2013). While even heathy adults show variability in performance (Hultsch, MacDonald, & Dixon, 2002), it is posited that greater variability in cognition reflects early signs of a brain under stress. Thus, greater IIV is associated with greater risk of cognitive decline and/or a subsequent dementia diagnosis (Anderson et al., 2016; Hultsch, MacDonald, & Dixon, 2002). Despite the straightforward nature of this hypothesis, differences in methodological and statistical approaches to studying variability in test performance in people at risk for dementia has meant that it is difficult to specify the conditions under which IIV is associated with later dementia. For example, two operationalisations of IIV are used commonly to define variability in cognitive test performance within individuals. These are, inconsistency (IIV-I), defined as variability within a cognitive test, and dispersion (IIV-D) defined as variability in the individual's performance across different tasks or domains. IIV-I operationalisations typically measure individual performance variability across multiple trials of a single task and is most commonly measured using variability on reaction time (RT) tasks (Hultsch, et al., 2000; Kochan et al., 2016). IIV-D operationalisations represent the variability an individual displays over separate cognitive tests within a domain (e.g. memory) or across different domains and is most commonly measured using the SD of z-transformed task scores (Holtzer,

Verghese, Wang, Hall & Lipton, 2008).While IIV-I and IIV-D both show promise in predicting AD-related decline or dementia diagnosis (Anderson et al., 2016; Bayer et al., 2014), the two approaches have seen little direct comparison in longitudinal research.

Understanding the association between IIV and later dementia has been complicated further by the different methods used to calculate IIV-I and IIV-D. For example, some authors compute standard deviations (ISD) of raw RTs on a single cognitive test (typically IIV-I) or on standardized performance scores across different tests (typically IIV-D). Another approach has been to utilize regression equations computed on raw scores (either IIV-I or IIV–D) and adjust for covariates such as age, gender, or mean test performance with the unexplained variance in such models defined as IIV(Anderson et al., 2016; Bayer et al., 2014b).

Intraindividual Variability (IIV)

IIV is a measure of an individual's ability to maintain globally consistent performance across trials and/or tasks of neuropsychological assessment measures. This can be contrasted with other performance measures used in neuropsychology which define cognitive performance in terms of the total, or average scores. The ability to maintain consistent performance reflected in IIV may be a sensitive marker of early neuropathological changes in AD and other dementias (Anderson, 2013; Kalin et al., 2014). Support for this hypothesis comes from studies reporting greater IIV predicts greater cognitive decline (Bielak, Hultsch, Strauss, MacDonald, & Hunter, 2010; Kliegel & Sliwinski Matthias, 2004), and conversion from cognitively normal ageing to later mild cognitive impairment (Anderson et al., 2016; Bayer et al., 2014). Furthermore, increasing IIV has been found to be associated with more direct indictors of AD pathological changes including 1) reduced white matter integrity (Head, Jackson, Balota, & Duchek, 2011; Mella, De Ribaupierre, Eagleson, & De Ribaupierre, 2013), 2) increased in the phosphorylated-tau/ Aβ-amyloid 42 ratio hallmark

pathological characteristic of AD (Patten, Fagan, & Kaufman, 2018) and 3), Apolipoprotein E (APOE) ε4 AD genetic risk status (Kalin et al., 2014; Tarnanas et al., 2015), albeit with moderate effect sizes in each case.

Differing approaches to IIV measurement

As stated above, variability in cognitive performance in individuals at risk for dementia has been operationalized in two main ways (IIV-I and IIV-D). While both approaches have shown an association with subsequent dementia, many theoretical models of cognitive IIV utilize these terms interchangeably. Thus while both approaches provide an index of cognitive variability, it remains possible that one is superior to the other (Anderson et al., 2016; Bayer et al., 2014). For example, direct comparisons of IIV-I and IIV-D in older adult cohorts, revealed that *only* IIV-I predicted APOEε4 status in a cognitively normal cohort, although with a small effect observed (Kalin et al., 2014; Tarnanas et al., 2015). Further, Tarnanas and colleagues (2015) suggest IIV-I may hold greater promise in identifying the early cognitive changes of prodromal AD. Specifically, IIV-I distinguished between cognitively normal (CN) and aMCI (MCI – amnestic type) while IIV-D distinguished between aMCI and AD, but not vice versa (CN vs aMCI). Christ and colleagues (2018) report RT based measures of IIV, common in IIV-I, are superior predictors of neurological impairment indexed by overall cognitive performance and memory predictors - compared to the total or average performance based measures typically used in IIV-D.

In addition to these operationalisations, different approaches to the measurement of IIV further complicates interpretation of the literature. A prominent difference in approach is whether estimates of variability are adjusted for covariates such as age or sex. Generally, this is achieved by regressing covariate(s) on IIV indices and then using the residuals in subsequent analyses. Some studies have adjusted for an individual's mean task performance in their calculation of IIV (Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000;

Lövdén, Li, Shing, & Lindenberger, 2007) while others have not (Salthouse, 2012; Tales et al., 2012). The decision to adjust for mean performance is based on the theoretical rationale that, particularly on RT tasks, mean RT should be considered when interpreting task variability since variability tends to increase with average RT increase (Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000). Similarly, some studies have adjusted their IIV measure for age (e.g. Anderson et al., 2016; Bielak et al., 2010) while others have not (Roalf et al., 2016; Salthouse & Soubelet, 2014). Those that control for age draw on evidence that older adults typically show greater IIV than younger adults (Hultsch, 2002) while those that do not control for age provide the rationale that IIV does not follow a consistent ageing pattern in later adulthood and is more likely to represent disease pathology (Roalf et al., 2016; Salthouse & Soubelet, 2014).

Summary and Purpose

The literature suggests greater IIV may be associated with cognitive decline and subsequent conversion to MCI/dementia (Bayer et al., 2014; Bielak et al., 2010; Anderson et al., 2016; Holtzer et al., 2008; Hultsch et al., 2002; Kliegel & Sliwinski Matthias, 2004). It is, however, unclear whether IIV, regardless of how it is operationalised or measured, is associated with cognitive decline and risk for conversion to dementia, or whether particular operationalisations are more useful. One way to improve our understanding of these different approaches and how they may exert influence on studies seeking to utilize cognitive variability to predict dementia is to conduct a meta-analyses of the extant literature that considers the extent to which cognitive variability is associated with subsequent dementia generally, as well as, how such estimates can be influenced by the different operational definitions and statistical approaches used in its computation. This study reports a meta-analysis of the association between IIV and subsequent cognitive decline or dementia diagnosis and compares IIV operationalisation and measurement approach as subgroups.

Method

Literature Search

A comprehensive electronic literature search was performed using PsycINFO, Embase, Medline, Scopus, and Google Scholar (extracted using Publish or Perish Software; Harzing., 2007) databases on the 11th of January 2021. The purpose of this search was to identify studies that have examined the association between IIV-I or IIV-D and subsequent cognitive decline or conversion to MCI/dementia. Search terms used were limited to "Within person or intra?individual variability IIV OR intra?individual OR individual differences OR cognitive variability OR dispersion" AND "Alzheimer* OR Alzheimer* disease OR dementia OR cognitive decline OR cognitive impairment OR mild cognitive impairment". Search results were limited to studies published in English. Searches included dissertations, theses and conference abstracts. Grey literature (including theses and conference abstracts) and reference lists of included studies were hand searched for studies that may have been missed in the electronic searches. These searches and subsequent screening steps are reported in line with PRISMA guidelines in Figure 1.

Studies were excluded at the title and abstract screening phase if they were irrelevant, duplicates, or clearly met exclusion criteria. The exclusion criteria included: reviews, studies that did not measure IIV of cognition (e.g. heart rate IIV), studies including samples with conditions or disorders other than dementia (e.g. Huntington's disease), cross-sectional designs, short follow-up (i.e. follow-up of fewer than 12 months), mean sample age less than 40 years (± 2SD), did not measure IIV (as defined by variance across or within cognitive tasks), non-cognitive/dementia outcome variable (e.g. fall risk), or duplicate sample (in which case the study with the largest sample size was selected). To confirm selection, 20% of titles and abstracts were screened by first and third authors, with 97% concordance. Any disagreement between raters was discussed by both and consensus reached.

At the full text screening phase, studies were included if they met the following inclusion criteria; 1) Examination of IIV-I or IIV-D and subsequent cognitive change/status, 2) report of an outcome measure providing quantifiable cognitive change effects on neuropsychological testing or diagnostic status. A further 10% of full text studies were screened by first and third authors with 100% concordance. Corresponding authors of studies missing key information where contacted (N = 4). No additional information that would allow for inclusion was provided and studies were not included in the analyses. These steps, as well as, the number of studies included or excluded at each step, are outlined in Figure 1.

The first author extracted the following data from each included study, 1) effect of the association between IIV performance and a change in cognitive performance on neuropsychological testing (correlation coefficient, beta weight) OR conversion to MCI/dementia, 2) IIV type (IIV-I or IIV-D), 3) whether the study examined cognitive decline or conversion to MCI/dementia, 4) whether the study adjusted for an individual's mean task performance, 5) whether the study controlled for other covariates e.g., age or gender, 6) sample size, 7) sample size characteristics including gender breakdown, baseline cognitive performance, and age information.

From 5948 studies initially identified, after duplicates were removed, title and abstract screening, full text screening and follow-up on studies with missing data (k = 4), 12 were included in the meta-analysis. These 12 studies (plus one additional study, noted below) are summarized in Table 1.

Independent Variable Classification

The independent variable in this meta-analysis was IIV, irrespective of IIV operationalisation (IIV-I or IIV-D) or measurement approach (i.e. regardless of whether or not the IIV measure controlled for mean performance or other demographic covariates). For studies reporting more than one measure of IIV, all IIV measures were included in the analyses, which were adjusted for multiple outcomes.

Subgroup Coding

IIV-I versus IIV-D: Included studies were reviewed to determine IIV operationalisation used. In total, 12 studies satisfied inclusion criteria for this meta-analysis. Of these, 7 estimated the relationship between IIV and cognitive decline or conversion to MCI/dementia using correlation coefficients (3 using IIV-D and 4 using IIV-I operationalisations) and 5 reported hazard ratios. Studies reporting hazard ratios could not be combined with studies reporting strength of association metrics (correlations/beta weights), since there is no accepted method of converting hazard ratios to correlation effects (Stare & Maucort-Boulch, 2016). Lead authors of studies reporting hazard ratio results were contacted to obtain raw data or alternative analysis results (e.g. odds ratios) to allow all 11 studies to be analysed together. One study (Kochan et al., 2016) provided odds ratio results. This study was incorporated with the seven association studies (now $N = 8$). The remaining four studies reporting hazard ratios were meta-analysed separately. Given recommendations for a minimum of four studies per group in categorical sub-group analyses (Fu et al., 2011), a previously excluded IIV-D study (Roalf et al., 2016- excluded due to a small overlapping sample with a study included in the Hazard ratio analysis; Anderson, Wahoske, Huber, Norton, Li, Koscik, Umucu, Johnson, Jones, Asthana, et al., 2016, but otherwise meeting

criteria) was subsequently included in the main analyses (now $N = 9$: 4 using IIV-D, 5 using IIV-I)[1].

Adjusting for Covariates versus Not Adjusting for Covariates

Included studies were also reviewed to determine whether they adjusted for mean performance or demographic covariates such as age in their IIV measurement approach. Studies were classified into those that controlled for mean performance versus those that did not, and into those that controlled for any other demographic covariates versus those that did not. Whilst, a subgroup analysis of these covariates was planned, this was not possible given the small number of studies available (Fu et al., 2011). As can be seen in Table 1, only two studies included in the general correlational analysis sought to control mean task performance, whilst only three sought to control other demographic covariates.

Outcome Variable Classification

Included studies were reviewed for outcome variable type, with studies using either 1) cognitive performance decline, or 2) conversion to MCI/dementia (i.e. CN, dementia, MCI, or MCI – amnestic). For a summary of outcome variable type see Table 1. Due to the small number of eligible studies, these outcome types were collapsed into a single outcome type representing dementia-related cognitive decline.

Data Analysis

Comprehensive Meta-Analysis v.3 (CMA; Borenstein, 2013) software using a random effects model was used to perform the meta-analysis. Two overall random effect analyses

[1] [1]This meant that paper reporting overlapping samples were included in both the correlational and hazard ratio meta-analyses. To examine the influence of this, the general correlational meta-analysis was re-run without the addition of the Roalf and colleagues (2016) study and the hazard ratio meta-analysis was also run without the Anderson and colleagues (2016) study. There was no substantive difference between the overall IIV effect estimates in the correlational or hazard ratio meta-analysis. Our preference was to report the larger, n = 4 analysis since this provides the best estimate available, especially given the Anderson study had the largest N. For results of these alternative analyses, see supplemental materials S2 and S3.

were conducted separating studies into correlational and hazard ratio analyses. Meta-analytic effects are reported as *r* with 95% confidence intervals ($N = 9$), where correlations of .10 were considered small effects, .20 medium, and .30 large (Gignac & Szodorai, 2016), or as hazard ratio results ($N = 4$). Follow-up meta-regression analyses (correlational studies only) were conducted to determine whether time to follow-up, baseline cognitive performance (Mini Mental State Examination (MMSE) scores were chosen as the most common assessment measure utilised by included studies), sex, or average age of participants had a significant influence on effect size. Heterogeneity was evaluated using the I^2 statistic which estimates the proportion of effect dispersion across studies representing real differences rather than random error (Lin, 2020). As I^2 is dependent on sample size (Von Hippel, 2015) heterogeneity was also inspected visually using forest plots, as well as, using Cochrane's Q and tau^2 statistics which indicate whether the observed variability is greater than that expected by chance.

Results

Overall effect

IIV (irrespective of operationalisation) was associated significantly with cognitive change (either cognitive decline or conversion to MCI/dementia) with a medium positive correlation of $r(7) = .20$, 95% CI = [.09, .31], $p < .001$ (see Figure 2). Follow-up meta-regression analyses of sex, baseline cognitive performance (MMSE score), time to follow-up, and average age of participants indicated no significant influence of sex $r(5) = -.00$, 95% CI = [-.01, .01], p = .849, or average age of participants $r(7) = -.002$, 95% CI = [-.01, .01], $p = .562$ on effect sizes. There was a significant influence of baseline cognitive performance $r(3) = -.08$, 95% CI = [-.15, -.01], p = 0.023 and time to follow-up $r(7) = -.003$, 95% CI = [-.01, -.00], $p = .002$ although these provided little explanation of IIV variance (both less than 1%).

Analysis of hazard ratio studies revealed baseline IIV (combining both IIV-I and IIV-D) significantly predicted cognitive change (either cognitive decline or conversion to MCI/dementia) with HR = 1.74, 95% CI = [1.02, 2.98], p = .044 (see Figure 3).

Heterogeneity

There was significant heterogeneity in the true IIV effect observed, I^2 = 88.11; $Q(8)$ = 67.29, p < .001; tau squared = 0.02 in the correlational analysis. As it is difficult to reliably interpret heterogeneity using I^2 when the number of included studies is small (Von Hippel, 2015), forest plots were also examined. These indicated significant heterogeneity was present with point estimates showing a range between r = .17 and r = .24. To determine whether analyses could proceed, a leave one out analysis was conducted (Wilcox, 2016). No change in the correlation effect size or significance value was noted with a medium positive correlation of $r(6)$ = .20 , 95% CI = [.09, .31], p <.001 suggesting that, despite significant heterogeneity, the overall effect size was robust at approximately .20. Given the small number of studies included in the HR analysis, heterogeneity analysis was not conducted.

Subgroup Analyses

Subgroup analysis, for the N = 9 correlation effects, revealed no difference in effect between IIV-I (r = .22, 95% CI = [.02, .41], N = 5) and IIV-D (r = .19, 95% CI = [.08, .29], N = 4) in their association with subsequent cognitive decline or conversion to MCI/dementia (Q = 3.41, p = .065), albeit this comparison should be interpreted with caution given the small number of studies (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Publication Bias

Funnel plots were examined to evaluate publication bias (see supplemental materials S1). The correlations were distributed asymmetrically with the smaller sample size studies shifting to the right, indicating bias. Egger‘s linear regression estimate = 4.51, p = .034. A trim-and-fill analysis suggested the possibility of two missing studies. Based on the inclusion

of these 'missing' correlations, the estimated adjusted correlation was r = .15, 95% CI = [.04, .25]. This suggests there may have been very slight overestimation of the overall correlation effect size. Analysis of bias was not appropriate for the hazard ratio analysis given the small number of included studies.

Discussion

The results of this meta-analysis indicate that IIV, regardless of operationalisation, was statistically significantly associated with subsequent cognitive decline or conversion to MCI/dementia. The average effect for the relationship between variability and subsequent cognitive decline or MCI/dementia was small with IIV explaining just 4% of the variance. There was no difference between IIV-I and IIV-D in their association with subsequent cognitive decline or MCI/dementia. While the absence of any difference may reflect the small number of studies that contributed data to these estimates (5 IIV-I studies and 4 IIV-D studies), this could mean that the effect (if any) is quite small. We had also planned to capture the effects of adjusting IIV measurement for covariates such as age or mean task performance, however, due to a small field this was not possible.

Cross-sectional evidence suggests that IIV-I is more strongly associated than IIV-D with genetic risk factors for AD (Kalin et al., 2014; Tarnanas et al., 2015), as well as, being better able to identify the earlier stages of AD (Christ, Combrinck, & Thomas, 2018a; Duchek et al., 2009; Kalin et al., 2014; Phillips, Rogers, Haworth, Bayer, & Tales, 2013). Despite this, no longitudinal empirical study has directly compared the association between later dementia and both IIV-I and IIV-D. Comparisons between studies that use different methods of operationalising IIV, as well as measurement approaches, may not reveal differences that would be more evident if IIV-I and IIV-D were compared within the same study. Moreover, the follow-up intervals for the studies reported here varied markedly. While

mean follow-up interval explained little variance in the overall effect of IIV on cognitive decline, it may have impacted the comparison of IIV-I and IIV-D. As can be seen in Table 2, mean follow-up for the IIV-I studies ranged from 30-156 months, *M* 64.8, whereas for IIV-D it ranged from 12-109 months, *M* 46.3. Further work is required to explore whether the lack of differences between IIV-I and IIV-D relates to the length of follow-up over which the effects are being evaluated.

Baseline cognitive performance explained some very small amount of variance in the overall effect of IIV on cognitive decline/dementia diagnosis. This is unsurprising as while IIV shows a unique pattern of change (Tractenberg & Pietrzak, 2011), it does indeed correlate with mean measures of performance (Nilam, Rabbitt, Brian, & John, 2005). Interpretation of this result is complicated by the baseline inclusion criteria utilised by each study with some studies opting for the baseline inclusion of healthy controls only while others chose to include participants classified as MCI at baseline. We need sufficient head to head comparisons of IIV (of either operationalisation) with more traditional neuropsychological measures such as mean scores, in order meta-analytically to confirm if IIV offers sensitivity to AD beyond that of mean performance. This will be useful in increasing our understanding of the true clinical utility of IIV and of alternative operationalisations.

The small number of studies currently available for inclusion in the meta-analysis, prevented our plan to explore the impact of adjusting IIV measurement for mean task performance or other demographic covariates such as age. More studies investigating the association between IIV and subsequent dementia are needed before meta-analytic investigation of these covariates can be conducted. It is recommended authors report both adjusted and unadjusted IIV separately to assist the field in determining which IIV operationalisation and measurement method offers the greatest association with cognitive decline in dementia.

Beyond measurement considerations, further empirical work would be helpful in evaluating whether IIV-I and IIV-D are conceptually similar or represent related but separate abilities. It could be argued that variability in RT within a test may reflect the degree to which an individual's processing resources are being taxed by that test. In contrast, when IIV is defined using variability in performance between tests, given that the tests likely differ in difficulty and nature, we do not know how much variability is 'normal'. Critically, the 'normal' level of variability may differ depending on the combination of tests being used, making IIV-D differentially sensitive to cognitive change when the degree of variability is marked. This is consistent with limited cross-sectional evidence that IIV-I may be more suited to detecting the subtle early changes seen in AD by predicting conversion to MCI, while IIV-D may be more suited to detecting later-stage decline by predicting conversion to AD (Christ, Combrinck, & Thomas, 2018b; Tractenberg & Pietrzak, 2011). The present study was precluded from examining the effect of different stages of disease progression due to the limited number of studies available (of those studies included in this meta-analysis, nine examined the association between IIV and conversion to MCI/dementia whilst four examined the association between IIV and potentially more subtle cognitive decline). Similarly, it would also be of interest to separate MCI, amnestic-MCI and dementia diagnosis outcomes to further investigate the relationship between IIV operationalisation and dementia progression.

Finally, we must also consider whether individual differences in effort, potentially as a result of depression, impact measures of IIV; though IIV-D perhaps more than IIV-I. This is because IIV-D compares performance between tests or tasks, meaning that those with depression or low effort may show more variability in task performance than those without, particularly if some tasks are more effortful or challenging than others (Freydefont, Golwitzer, & Oettingen., 2016). Whether IIV-I is impacted by effort/depression may depend on how it is measured. Some studies (e.g. Lovden et al., 2007) analyse RTs for correct trials

only, or use trials with RTs in a range suggestive of appropriate attention to task i.e. not too slow or too fast (Tales et al., 2012). This is done in an attempt to remove the impact of poor effort on performance. Some studies also choose to control for mean RTs (Lovden et al., 2007), which would control for slowing due to reduced effort. Given that there were no differences between IIV-I and IIV-D in their ability to predict conversion to dementia, and that some studies excluded depression psychiatric disorders such as depression (Bayer et al., 2014; Bielak et al., 2010) it seems unlikely that depression is what drives conversion to dementia. Further to this, there is evidence to suggest IIV follows an inverted U shape across the lifespan (Hultsch, MacDonald, & Dixon, 2002), suggesting IIV exists independently of (but not necessarily unaffected by) effort and/ or depression. It is possible, however, that IIV is affected by individual differences in effort more generally, and future studies should explore effort as a covariate of IIV.

Conclusion

Cognitive IIV appears to hold a statistically significant association with subsequent cognitive decline with a medium effect size noted. This is consistent with a growing body of research suggesting cognitive variability is a promising indicator of early brain changes in dementia. Unfortunately, the evidence does yet allow us to conclude whether IIV-I or IIV-D differ in their association with cognitive decline or conversion to MCI or dementia. Nor are we yet able to advise whether varying approaches to IIV measurement (e.g. adjusting for mean performance or other demographic covariates) are poorer or stronger indicators of incipient cognitive decline. Overall, cognitive IIV, including both IIV-I and IIV-D operationalisations, appears to hold a significant association with cognitive decline and may offer a novel indicator of incipient dementia in both clinical and research settings.

References

Anderson, E. (2013). Cognitive variability on neuropsychological testing and hippocampal volume loss as predictors of incident Alzheimer's disease in a geriatric population: Analysis from the Alzheimer's disease neuroimaging initiative. *Dissertation Abstracts International: Section B: The Sciences and Engineering*. Retrieved from http://search.proquest.com/openview/9524727d72e4b15cb8df76aee1d628ff/1?pq-origsite=gscholar&cbl=18750&diss=y

Anderson, E. D., Wahoske, M., Huber, M., Norton, D., Li, Z., Koscik, R. L., … Asthana, S. (2016). Cognitive variability-A marker for incident MCI and AD: An analysis for the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *4*, 47–55. https://doi.org/http://dx.doi.org/10.1016/j.dadm.2016.05.003

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 Alzheimer's Disease*, *40*(1 PG-177–189), 177–189. https://doi.org/10.3233/JAD-131934

Bielak, A.A.M., Hultsch, D. F., Strauss, E., MacDonald, S. W. S., & Hunter, M. A. (2010). Intraindividual Variability in Reaction Time Predicts Cognitive Outcomes 5 Years Later. *Neuropsychology*, *24*(6), 731–741. https://doi.org/10.1037/a0019802

Borenstein, M., Hedges, L. V, Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Retrieved from www.wiley.com.

Borenstein M, H. L. (2005). Comprehensive Meta-analysis Version 2, Biostat.

Christ, B. U., Combrinck, M. I., & Thomas, K. G. F. (2018). Both Reaction Time and

Accuracy Measures of Intraindividual Variability Predict Cognitive Performance in Alzheimer's Disease. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00124

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*. Retrieved from http://ezproxy.library.uwa.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc6&AN=2009-20255-007

Freydefont, L., Gollwitzer, P. M., & Oettingen, G. (2016). Goal striving strategies and effort mobilization: When implementation intentions reduce effort-related cardiac activity during task performance. *International Journal of Psychophysiology*, *107*, 44-53.

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., … Trikalinos, T. A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of Clinical Epidemiology*. https://doi.org/10.1016/j.jclinepi.2010.08.010

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Head, D., Jackson, J., Balota, D., & Duchek, J. (2011). White matter integrityand intraindividual variability in aging and early-stage alzheimer disease. *Alzheimer's and Dementia*, *1*, S736. https://doi.org/http://dx.doi.org/10.1016/j.jalz.2011.05.2119

Holtzer, R., Jacobs, S., & Demetriou, E. (2020). Intraindividual variability in verbal fluency performance is moderated by and predictive of mild cognitive impairments. *Neuropsychology*, *34*(1), 31.

Holtzer, R., Verghese, J., Wang, C., Hall, C. B., & Lipton, R. B. (2008). Within-person across-neuropsychological test variability and incident dementia. *JAMA - Journal of the American Medical Association*, *300*(7), 823–830. https://doi.org/10.1001/jama.300.7.823

Hultsch, D.F., MacDonald, S. W. S., Hunter, M. A., & Levy-Bencheton, J. (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. https://doi.org/http://dx.doi.org/10.1037/0894-4105.14.4.588

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). https://doi.org/10.1093/geronb/57.2.P101

Kalin, A. M., Pfluger, M., Gietl, A. F., Riese, F., Jancke, L., Nitsch, R. M., & Hock, C. (2014). Intraindividual variability across cognitive tasks as a potential marker for prodromal Alzheimer's disease. *Frontiers in Aging Neuroscience*, *6*(JUL), 147. https://doi.org/http://dx.doi.org/10.3389/fnagi.2014.00147

Kliegel, M., & Sliwinski Matthias; ORCID: http://orcid.org/0000-0002-2001-2522, M. A. I.-O. http://orcid. org/Kliege. (2004). MMSE Cross-Domain Variability Predicts Cognitive Decline in Centenarians. *Gerontology*, *50*(1), 39–43. https://doi.org/http://dx.doi.org/10.1159/000074388

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. *The American Journal of Geriatric Psychiatry*, *24*(3), 221–231. https://doi.org/http://dx.doi.org/10.1016/j.jagp.2015.12.005

Koscik, R., Berman, S., Clark, S., Mueller, K., Okonkwo, O., Gleason, C., … Johnson, S. (2016). Intraindividual cognitive variability in middle age predicts cognitive impairment 8–10 years later: results from the Wisconsin Registry for Alzheimer's Prevention. *Journal of the International Neuropsychological Society*, *22*, 1016–1025. Retrieved from https://www.cambridge.org/core/journals/journal-of-the-international-neuropsychological-society/article/intraindividual-cognitive-variability-in-middle-age-predicts-cognitive-impairment-810-years-later-results-from-the-wisconsin-registry-for-alzheimers-pr

Lin, L. (2020). Comparison of four heterogeneity measures for meta-analysis. *Journal of Evaluation in Clinical Practice*, *26*(1), 376–384. https://doi.org/10.1111/jep.13159

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*(12), 2827–2838. https://doi.org/10.1016/j.neuropsychologia.2007.05.005

Mella, N., De Ribaupierre, S., Eagleson, R., & De Ribaupierre, A. (2013). Cognitive intraindividual variability and white matter integrity in aging. *The Scientific World Journal*, *2013*. https://doi.org/10.1155/2013/350623

Nilam, R., Rabbitt, P., Brian, S., & John, N. (2005). Cognitive Performance Inconsistency : Intraindividual Change. *Psychology and Ageing*, *20*(4), 623–633. https://doi.org/10.1037/0882-7974.20.4.623

Patten, R. Van, Fagan, A. M., & Kaufman, D. A. S. (2018). Differential Cued-Stroop Performance in Cognitively Asymptomatic Older Adults with Biomarker-Identified Risk for Alzheimer's Disease: A Pilot Study. *Current Alzheimer Research*, *15*(9), 820–827. https://doi.org/https://dx.doi.org/10.2174/1567205015666180404170359

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 [Electronic Resource]*, *8*(6). https://doi.org/https://dx.doi.org/10.1371/journal.pone.0065712

Roalf, D. R., Quarmley, M., Mechanic-Hamilton, D., Wolk, D. A., Arnold, S. E., & Moberg, P. J. (2016). Within-individual variability: An index for subtle change in neurocognition in mild cognitive impairment. *Journal of Alzheimer's Disease*. Retrieved from http://ezproxy.library.uwa.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc13&AN=2016-45408-029

Salthouse, T A. (2012). Psychometric Properties of Within-Person Across-Session Variability in Accuracy of Cognitive Performance. *Assessment*, *19*(4), 494–501. https://doi.org/10.1177/1073191112438744

Salthouse, Timothy A, & Soubelet, A. (2014). Heterogeneous ability profiles may be a unique indicator of impending cognitive decline. *Neuropsychology*, *28*(5 PG-812–8), 812–818. https://doi.org/https://dx.doi.org/10.1037/neu0000100

Stare, J., & Maucort-Boulch, D. (2016). *Odds Ratio, Hazard Ratio and Relative Risk. Metodološki zvezki* (Vol. 13).

Tales, A., Leonards, U., Bompas, A., Snowden, R. J., Philips, M., Porter, G., … Bayer, A. (2012). Intra-Individual reaction time variability in amnestic mild cognitive impairment: A precursor to dementia? *Journal of Alzheimer's Disease*, *32*(2), 457–466. https://doi.org/10.3233/JAD-2012-120505

Tarnanas, I., Papagiannopoulos, S., Kazis, D., Wiederhold, M., Widerhold, B., & Tsolaki, M. (2015). Reliability of a novel serious game using dual-task gait profiles to early characterize aMCI. *Frontiers in Aging Neuroscience*, *07*.

https://doi.org/10.3389/fnagi.2015.00050

Tractenberg, R. E., & Pietrzak, R. H. (2011). Intra-individual variability in Alzheimer's disease and cognitive aging: definitions, context, and effect sizes. *PLoS ONE [Electronic Resource]*, *6*(4), 19. https://doi.org/https://dx.doi.org/10.1371/journal.pone.0016973

Vaughan, L., Leng, I., Dagenbach, D., Resnick, S. M., Rapp, S. R., Jennings, J. M., … Espeland, M. A. (2013). Intraindividual variability in domain-specific cognition and risk of mild cognitive impairment and dementia. *Current Gerontology and Geriatrics Research*, *2013*, 495793. https://doi.org/https://dx.doi.org/10.1155/2013/495793

Von Hippel, P. T. (2015). The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Medical Research Methodology*, *15*(1). https://doi.org/10.1186/s12874-015-0024-z

Wilcox, R. R. (2016). *Introduction to Robust Estimation and Hypothesis Testing: 4th Edition*. *Introduction to Robust Estimation and Hypothesis Testing: 4th Edition*. Retrieved from www.elsevierdirect.com

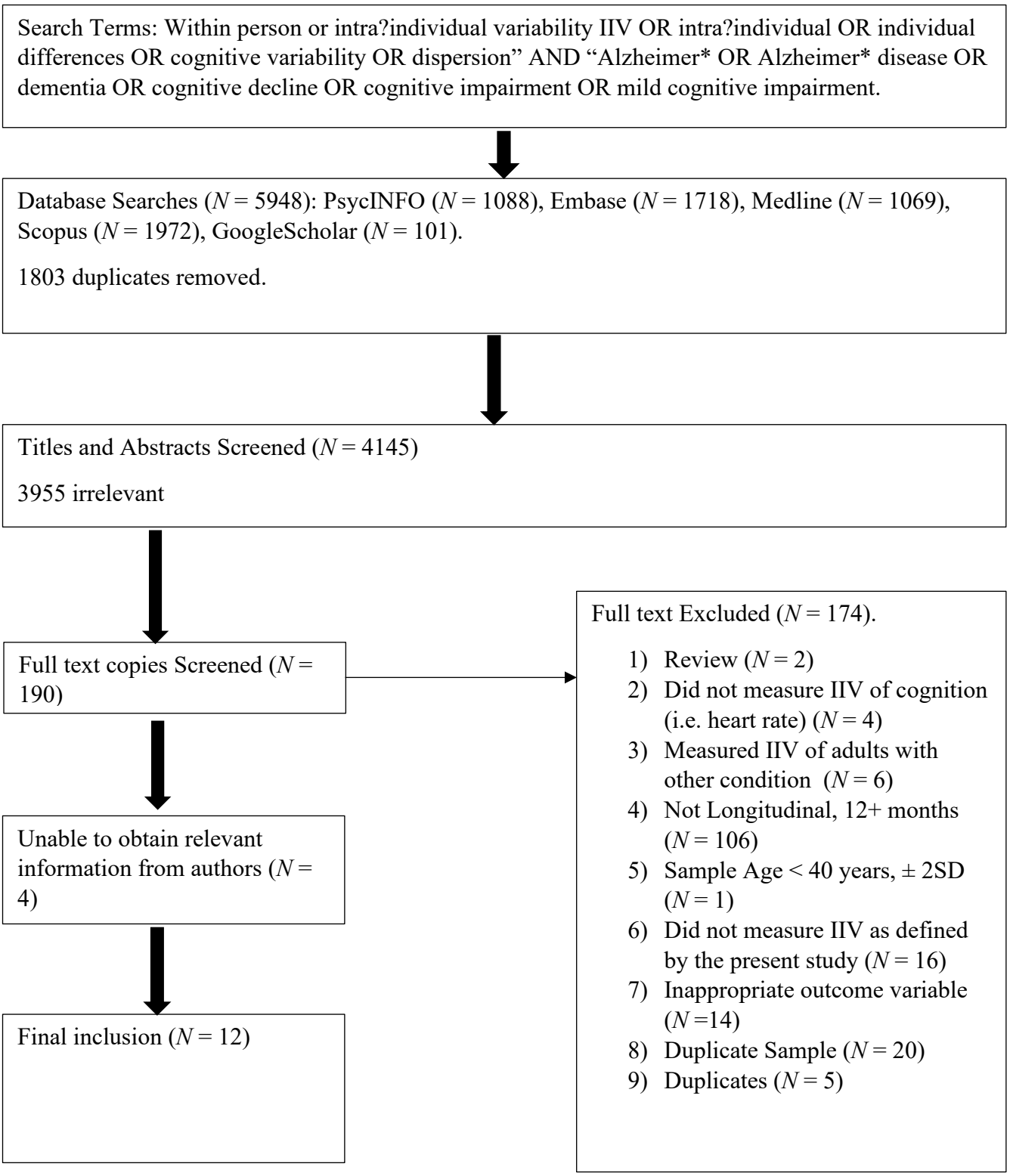

Figure 1. Systematic Search and Screening Results (PRISMA chart).

Table 1

Summary of included studies.

Study	Sample Size	Mean Sample Age (years)	Sample Sex % Female	IIV Sub-type	Cognitive Tasks used	Mean Adjusted	Other Covariates Adjusted	Outcome Measure	Author's Conclusion	Mean Follow-Up (months)
General Correlation Analysis										
Bayer et al., 2014	76	72.9	51	IIV-I	Posner exogenous cueing paradigm.	No	No	Conversion to MCI/dementia	Higher IIV at baseline was associated with development of dementia.	30
Bielak et al., 2010	212	74.3	68	IIV-I	Finger tapping, four choice reaction time, four choice reaction time 1 back, shape, colour and task switching.	No	Yes (Age, practise effects)	Conversion to MCI/dementia	Greater IIV was associated with greater likelihood of being in the maladaptive group (Cognitive impairment, no dementia).	60
Lovden et al., 2007	447	84.1	Not Reported	IIV-I	Identical pictures test.	Yes	Yes (Time to death, age, and suspected dementia).	Cognitive decline	High IIV signals impending cognitive decline.	156
[B] Roalf et al., 2016	819	74	42	IIV-D	Rey Auditory Verbal Learning Test, Wechsler Memory Scale-Revised, Logical Digit Span (forward and backward), Trail Making (Part A and B),	No	No	Conversion to MCI/dementia	Variability at baseline was higher in individuals transitioning from MCI to AD.	12

					Digit Symbol Substitution, Semantic word list generation, Boston Naming, Alzheimer's disease Assessment Scale – Cognition subscale, Clock drawing.					
Salthouse et al., 2014	352	73.1	57	IIV-D	Matrix Reasoning, Shipley Abstraction, Letter Sets, Spatial Relations, Paper Folding, Form Boards; Word Recall, Paired Associates, Logical Memory, Digit Symbol, Pattern Comparison, Letter Comparison.	No	No	Cognitive decline	Initial IIV was greater for those who experienced most longitudinal change.	33
Tales et al., 2012	39	72.9	51	IIV-I	Exogenous target detection cueing paradigm.	No	No	Conversion to MCI/dementia	IIV differentiated those with MCI who converted to dementia from non-convertors.	30
Kliegel et al., 2004	91	100.2	Not Reported	IIV-D	MMSE.	Yes	No	Cognitive decline	IIV predicted cognitive decline better than mean performance.	18
Kochan et al., 2016	861	78.7	55	IIV-I	Simple and complex RT tasks.	No	No	Conversion to MCI/dementia	IIV independently predicted time to dementia.	48
Koscik et al., 2016	684	53.6	70	IIV-D	Rey Auditory Verbal Learning Task, Trail Making Test (A & B), Wide Range Achievement Test-3rd edition	No	Yes (gender, literacy, family history of AD, *APOE* ε4 carrier, baseline age, follow-up time)	Conversion to MCI/dementia	IIV predicted subsequent impairment. Prediction was weaker than mean memory and executive function scores.	109

Hazard ratio Analysis										
[A] Anderson et al, 2016	1324	73.7	44	IIV-D	Rey Auditory Verbal Learning Test, Total of learning trials, Rey American National Adult Reading Test, Trail Making (A & B).	Yes	Yes (age, education, *APOE* ε4)	Conversion to MCI/dementia	IIV was associated with time to cognitive status change.	31
Holtzer et al., 2008	897	78.6	60	IIV-D	The Free and Cued Selective Reminding Test, WAIS-R Vocab and Digit symbol substitution.	Yes	Yes (sex, education, medical illness)	Conversion to MCI/dementia	IIV was associated with development of dementia independent of mean test performance.	40
Holtzer et al., 2020	344	75.89	55	IIV-I	Semantic and Letter Fluency.	No	No	Conversion to MCI/dementia	Baseline IIV on semantic, but not letter fluency tasks, predicted MCI.	35.4
Vaughan et al., 2013	2305	74.0	100	IIV-D	Primary mental abilities test of verbal knowledge, Benton Visual Retention Test, California Verbal Learning Test, Digit span forward and backward, Card rotations test, letter and semantic fluency, finger tapping,	No	No	Conversion to MCI/dementia	IIV significantly predicted dementia.	64

Note. Mean controlled = mean performance used as a covariate of IIV (usually through calculating the Coefficient of variation); covariates controlled = demographic information used as a covariate of IIV (usually through the calculation of residuals); Cognitive decline = decline in cognitive performance on neuropsychological measures over visits; Conversion to MCI/dementia = Clinical Diagnosis of CN, MCI, MCI (amnestic), AD or dementia. Abbreviations: AD, Alzheimer's disease, *APOE,* Apolipoprotein E gene; CN, Cognitively normal; IIV-I, Intraindividual Variability- Inconsistency; IIV-D, Intraindividual Variability – Dispersion; MCI, Mild Cognitive impairment; RT, Reaction time. [B] Roalf et al ., (2016) contains an overlapping sample to that reported in [A] Anderson et al., (2016) and was initially excluded. As noted, it was included to allow for sub group comparison in the correlational meta-analysis.

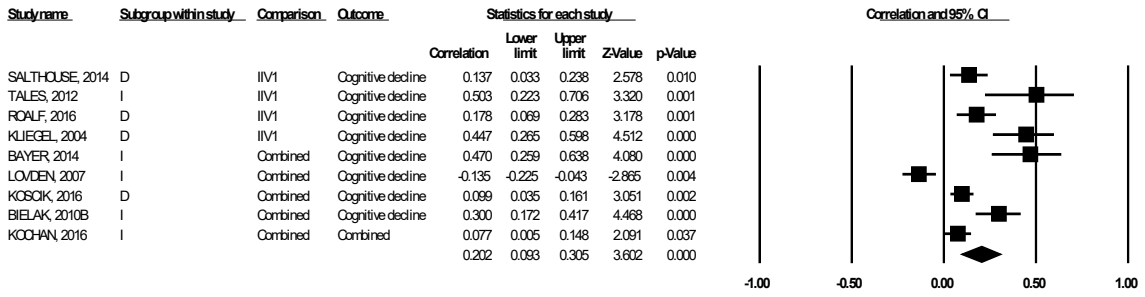

Figure 2. Forest plot of observed random effects correlation coefficients for IIV predicting cognitive change (cognitive decline or conversion to MCI/dementia). Squares represent study effect size with horizontal solid lines representing 95% CI. Diamond represents overall effect size. Combined = study reporting more than one IIV correlation (e.g. IIV calculated using multiple tasks).

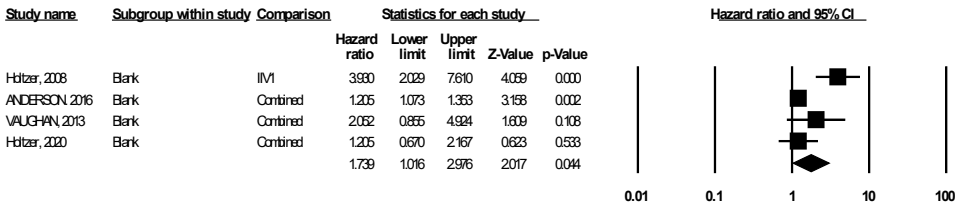

Figure 3. Forest plot of observed random effects hazard ratios for IIV predicting cognitive change (cognitive decline or conversion to MCI/dementia). Squares represent study effect size with solid line representing 95% CI. Diamond represents overall effect size. Combined = study reporting more than one IIV correlation (e.g. IIV calculated using multiple tasks).

Supplemental Material

S1 – Funnel Plots

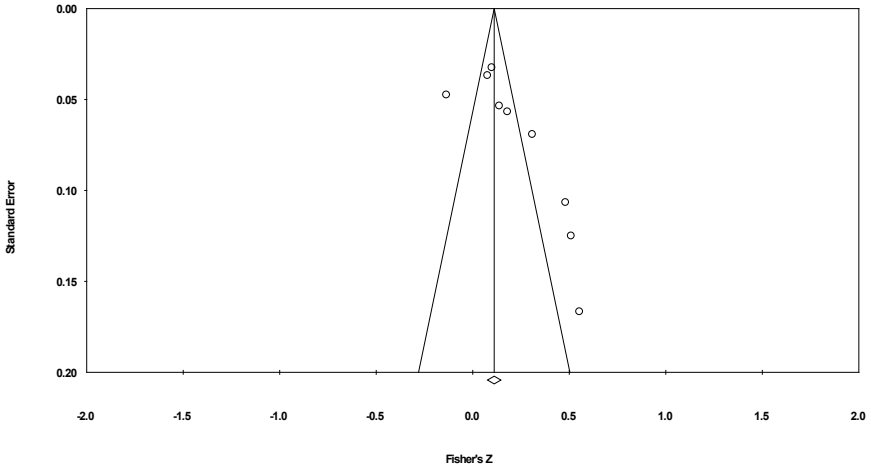

Figure 1a. Funnel plot based on observed correlations for IIV in predicting cognitive decline/subsequent diagnosis of dementia from general correlational analysis. White circles represent observed correlations from included studies.

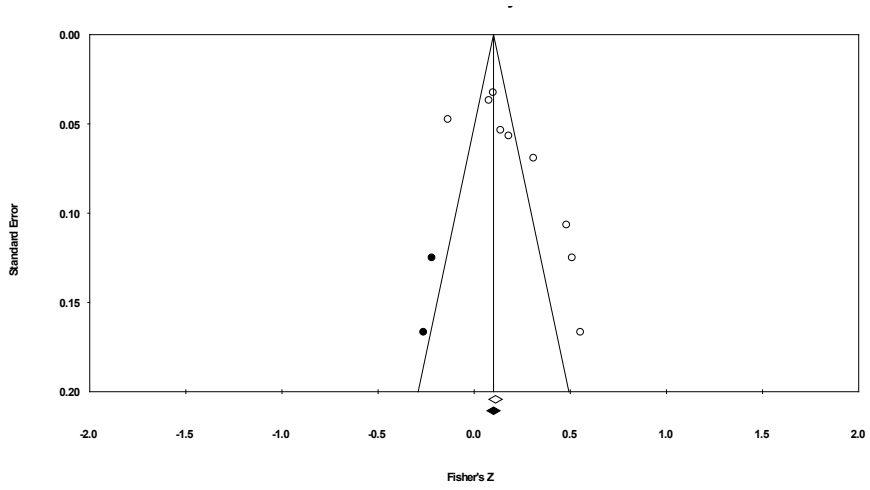

Figure 1b. Funnel plot based on observed and estimated correlations for IIV in predicting cognitive decline/subsequent diagnosis of dementia from general correlational analysis. White circles represent observed correlations from included studies, black circles represent estimated correlations from proposed unpublished studies.

S2 – Forrest Plots Correlation Meta-analysis (*N* = 8)

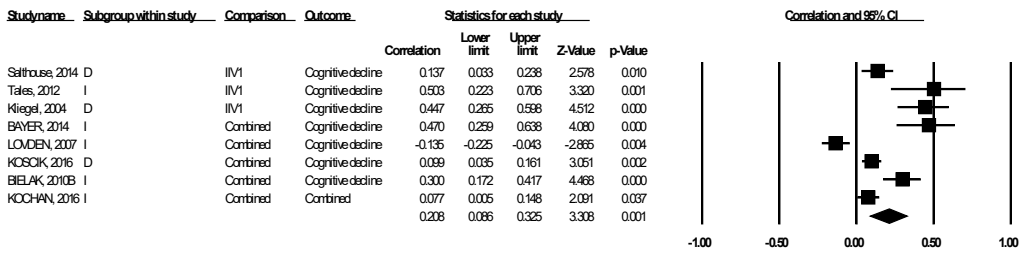

Figure 1. Forest plot of observed random effects correlation coefficients for IIV and subsequent cognitive change (cognitive decline or conversion to MCI/dementia), without the inclusion of the Roalf and colleagues (2016) study. Squares represent study effect size with horizontal solid lines representing 95% CI. Diamond represents overall effect size. Combined = study reporting more than one IIV correlation (e.g. IIV calculated using multiple tasks).

S2 – Forrest Plot Hazard Ratio Meta-analysis (*N* = 3)

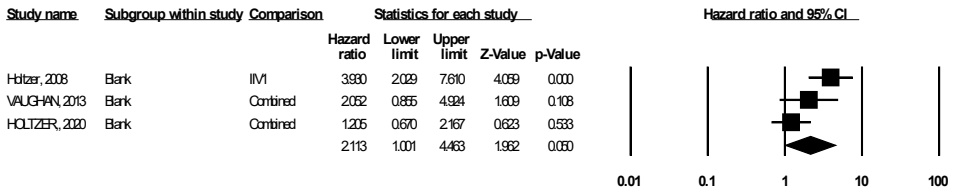

Figure 1. Forest plot of observed random effects hazard ratios for IIV predicting cognitive change (cognitive decline or conversion to MCI/dementia), without the inclusion of the Anderson and colleagues (2016) study. Squares represent study effect size with solid line representing 95% CI. Diamond represents overall effect size. Combined = study reporting more than one IIV correlation (e.g. IIV calculated using multiple tasks).