Using Response Time Data from Social Science Surveys to Model Cognition and Cognitive Decline

Seth Sanders
Professor of Economics, Duke University and Senior Fellow at NORC

William Dale
Section Chief of Geriatrics & Palliative Medicine, University of Chicago Medicine

Henrik Olsson
Associate Professor of Psychology, University of Warwick & External Faculty, Santa Fe Institute

Lynne Schofield
Associate Professor of Statistics, Swarthmore College

L. Philip Schumm
Department of Health Studies, University of Chicago

Linda Waite
Lucy Flower Professor in Urban Sociology, University of Chicago

November 23, 2017
Preliminary and Incomplete

Abstract: Social science researchers are increasingly interested in cognitive aging and its relationship to other life events. Most longitudinal datasets have no or limited direct measures of cognition. Using the National Social, Health and Aging Project (NSHAP) we show that the time it takes to answer questions measuring cognition is highly correlated with measured levels and declines in cognition. These measures are also highly correlated with 5 year mortality. Data on the time to answer questions is routinely captured as a by-product of computer assisted interviewing yet it is rarely used by the social science research community. Our results suggest a large amount of useful information is likely contained within most social science surveys that has to date gone unused and may be useful for modeling the aging process.

This work is supported by "Using Response Time Data from Social Science Surveys to Model Behavioral Outcomes,"(1R21HD088618), "The National Social Live, Health and Aging Project" (5R01AG021487) "Using Response Time Data from Social Science Surveys to Model Cognition and Early Alzheimer's Disease" (and P30 AG034424-S1)

1. Introduction

While neuropsychological testing remains the gold standard for assessing “pathological” cognitive changes in the older adult population, the interest in and necessity of measuring cognition as part of population-based studies are increasing (Nathan, Wilkinson, Stammers, and Low 2001). The Health and Retirement Study (HRS) has contributed longitudinal and cross-sectional information on cognition since 1992. It initially assessed four cognitive aspects—memory, abstract reasoning, self-rated cognitive functioning, and “functioning in cognitively demanding activities of daily living” (Wallace & Herzog 1995). More recently, HRS has expanded cognitive testing to include additional domains such as attention and verbal fluency as part of the Telephone Interview for Cognitive Status (TICS). A substudy of HRS, the Aging, Demographics, and Memory Study (ADAMS) incorporated a comprehensive neuropsychological assessment of a selected group to better differentiate and understand normal cognition from cognitive impairment, no dementia, and dementia (Langa et al. 2005).

Internationally, studies of aging have also begun to incorporate cognitive measures. For instance, the English Longitudinal Study of Health and Aging (ELSA), a large study of community-dwelling individuals in the United Kingdom, assessed time orientation, immediate and delayed verbal recall, prospective memory, verbal fluency, numerical ability, cognitive speed, and attention (Llewellyn, Lang, Langa, and Huppert 2008). The Canadian Study of Health and Aging (CSHA) incorporated the modified Mini-Mental State Exam (MMSE) (Teng and Chui 1987). The Swedish Betula study (Nilsson et al. 2004) and the Berlin Aging Study (Baltes and Mayer 1999) include several cognitive neuroscience-based functioning measures including episodic, semantic, and priming tasks.

Until recently U.S. and international studies continued to overlook important components of cognitive functioning, namely visuo-construction skills and executive function. A major recent advance is the development of a survey based-version of the Montreal Cognitive

Assessment (MoCA) that was first incorporated in wave 2 of the National Social, Health and Aging Project (NSHAP) and repeated five years later in wave 3 of the survey (Shega, et al. 2014; Kotwal et al. 2014). The MoCA includes components that measure visuo-construction skills and executive function. Importantly, the original MoCA test validation study (Nasreddine et al. 2005) has shown it to be a promising tool for detecting mild cognitive impairment (MCI) and the early stages of AD. The sensitivity and specificity of the MoCA for detecting AD was 100% and 87% respectively. Having two measures in time of the MoCA allows us to study both cognitive functioning in the cross-section and changes in cognitive functioning over time.

To date, almost all of studies have relied on test-based measures of cognitive assessment that use some function of the number of questions answered correctly to form a cognitive assessment score (for exception, see Tse et al. (2010) and Lövden et al. (2007)). In an entirely separate line of work, psychologists and neuroscientists have used the time it takes to answer cognitive challenges, usually referred to as response time (RT), to measure a subject's level of cognition. Specifically psychologists and neuroscientists have developed sequential sampling models that formalize how neural circuitry operates as human decision making occurs, for example the Drift Diffusion Model (DDM) (Radcliff 1978; Smith 2000). One of the more surprising findings from this literature is that an individual's ability to think in highly complex and abstract forms is related to speed in tasks as simple as "press the lighted button." Simple RT tasks appear to have predictive power for performance on much more elaborate tasks

Some studies relate parameters such as the variability of RT on neuroscience tasks either to changes over time in mean RT on these or related tasks, or explicitly to dementia. Perhaps the most related work to ours is Tse et al. (2010). They characterize the RT distribution to three computerized tests: the Stroop test, the Simon test, and the switching test; and they investigate how the RT distribution differs between young adults, old adults, and people with very mild dementia of the Alzheimer Type (DAT) as assessed on the clinical dementia rating (CDR). They show strong evidence that both the error rates are higher and the RT longer for the DAT group. They do not ask if RT might predict future changes in the CDR, and they examine individuals who, while classified as DAT, are of near-normal functioning. (On the MMSE, the DAT sample had a mean MMSE score of 26.5, higher than the traditional cutoff for MCI of 24). Ratcliff et al. (2010) analyze three two-choice tasks: numerosity discrimination, recognition memory, and lexical decision, and relate the RT distribution on these to age and IQ but not to standard measures of cognitive function. Using the 13-year longitudinal Berlin Aging Study, Lövden et al. (2007) analyze data from two neuroscience RT tasks and show that variance in RT early in the survey predicts increases in RT the next several years. But this study does not relate these neuroscience tasks to standard cognitive assessments. In addition, most of these studies have relatively small sample sizes; the laboratory studies typically have fewer than 150 subjects and the panel surveys typically fewer than 500 respondents, which limits the ability to understand covariates associated with cognition or changes in cognition.

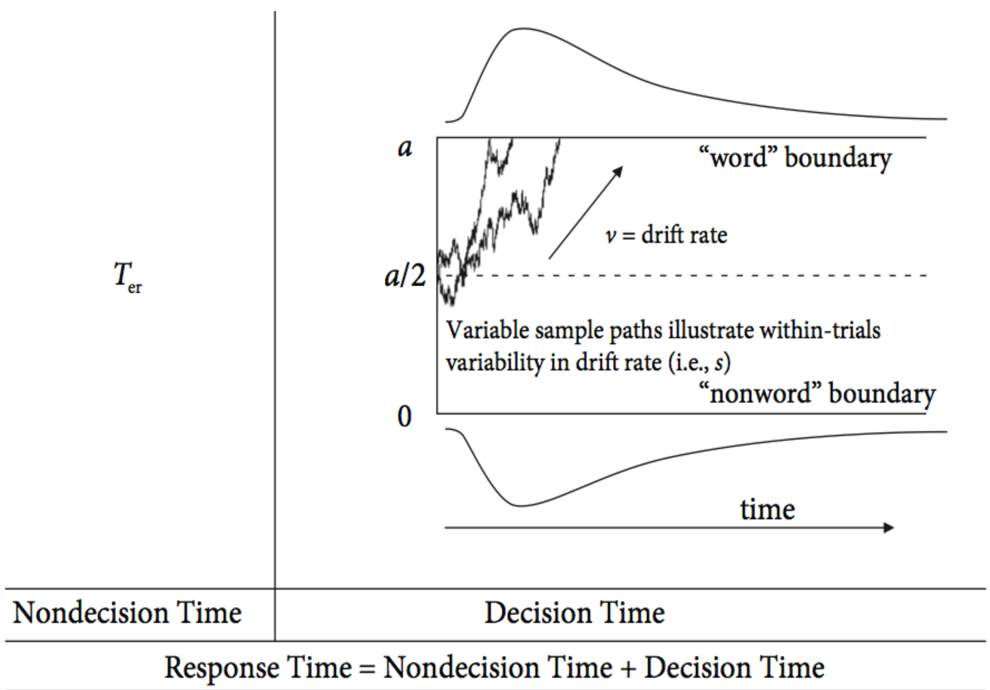
This paper brings together these two approaches within a large longitudinal population survey. As in neuroscience, we are interested in RT measures of cognitive functioning, but our goal is to investigate whether useful information is available in RT data that are collected as a by-product of normal survey-taking operations. The NSHAP implements a survey version of the MoCA, the MoCA-SA, and a host of survey questions within a Computer Assisted Personal Interviewing (CAPI) system that records time stamps for each question asked. From these time stamps, first we investigate whether there is a relationship between the speed in answering the MoCA-SA and the MoCA-SA score itself. Then we investigate whether the speed in answering the MoCA-SA is predictive of changes in cognitive capacity as measured by changes in the MoCA-SA score over the five years between waves 2 and 3 of the survey. Finally, we investigate whether RT is predictive of mortality between waves 2 and 3. Mortality has been closely associated with levels and decline of cognition.

Our context is quite different from the laboratory environment. In large population-representative studies, many factors influence RT of survey participants, from item-level characteristics such as number of clauses and words in a question; respondent-level characteristics such as age and education; mode characteristics such as whether the survey is self or interviewer administered; and for interviewer-administered surveys, some interviewer characteristics such as age and experience (Couper and Kreuter 2013; Yan and Tourangeau 2008). We need to account for factors about interviewers. We do this by controlling by including interviewer fixed effects. A second challenge is that our cognitive tasks are the byproduct of normal interviewing procedures. That is, our tasks were not designed to have RT as a primary outcome. Many cognitive tasks in neuroscience are designed to have a forced answer over a limited period of time, typically less than two seconds. Often survey questions take longer to answer and "don't know" is often one potential response. Because of these factors, we use models in cognitive neuroscience as a guide, but we remain open to processes generating RT on surveys that are different from on laboratory-administered tasks. Our primary goal, however, remains to capture processes of RT and relate them to traditional measures of cognition and cognitive change.

2. Neuroscience Theory of Cognitive Functioning and Response Time

Psychologists and neuroscientists have, over the last twenty-five years, developed sequential sampling models that formalize how neural circuitry operates as human decision making occurs. One example is the Drift Diffusion Model (DDM) (Ratcliff 1978; Smith 2000). Figure 1 displays a simple model of the DDM found in Wagenmakers et al. 2007. The DDM assumes decisions are made by a noisy process that accumulates information over time until a boundary level of evidence is reached that one choice is best and a decision is initiated. For example, Figure 1 describes an experiment in which "words" are flashed on a computer screen; some "words" are real while others have had vowels changed so that they are not real English words. The subject is asked to decide whether the word is real or not. The theory is that there is a non-decision time that subjects take to understand the task they are trying to solve (for example reading and understanding instructions) and then a time it takes to solve the task once the task is understood. The rate of accumulation of evidence, known as the drift rate, is faster if the relative difficulty of a choice is low. For example, words used frequently are easy to identify and can be easily distinguished from manipulated words (e.g, pig vs pyg) while lower frequency words are more difficult to identify and distinguish from manipulated words (polygon vs. poligon). The

Figure 1: EZ Drift Diffusion Modell (From Wagenmakers, et al., 2007)



drift rate is thought to vary across individuals. For any given drift rate, the model is able to describe the relationship between the relative value of choices, the choice made, and the speed at which the choice is made. Another task often explored is to display two lines on a computer screen and to ask the subject to choose the line that is longer. The model predicts that given the subject's specific processing speed, the greater the difference in the lengths of the lines: (a) the more likely the subject will pick the longer line correctly; (b) the faster the subject will pick the longer line; and (c) it will take longer when the shorter line is mistakenly chosen than when the longer line is correctly chosen. In addition, the model predicts an interaction between processing speed and task difficulty. Specifically, easy tasks can be answered correctly and quickly by most subjects while difficult tasks are answered faster and more accurately by subjects with higher cognitive capacity. A key insight of this and other neuroscience models is that multiple outcomes are determined by the relative difficulty of tasks and by the subject's own processing speed.

Radcliff (1979) suggests estimating response time distributions using an ex-Gaussian function which is achieved by maximizing the following likelihood function across the Q items that each with a RT recorded:

$$logL(\mu_i, \sigma_i, \tau_i | t_{i1}, \dots, t_{iQ}) = -\sum_{q=1}^{Q} ln\left(\frac{1}{\tau_i} exp\left(\frac{\mu_i}{\tau_i} + \frac{{\sigma_i}^2}{2{\tau_i}^2} - \frac{t_{iq}}{\tau_i}\right) \Phi\left(\frac{t_{iq} - \mu_i - ({\sigma_i}^2/\tau_i)}{\sigma_i}\right)\right) \quad (1)$$

If questions differ in their difficulty, we measure each individual's response time on a specific question q as a Z-score, that is $t_{iq} = \left(t_{iq}^* - \mu_q^*\right) / \sigma_q^{*2}$ where t_{iq}^* denotes actual response time for respondent i on question q and μ_q^* and σ_q^* are the mean RT and standard deviation of RT for question q. μ_i, σ_i and τ_i then characterize respondent *i's* RT distribution in terms of time to set up decision making and time to respond. Radcliff's work is motivated by the DDM model of RT where two mental processes are co-integrated to lead to the RT of a task. The nondecision time is modeled as an exponential process ("ex") while the Normal (Gaussian) component is seen as modeling the time it takes the sensory process to accumulate information on choices and the time required to physically respond with a decision, the decision time (hence the term "ex-Gaussian"). While the theoretical proposition that the RT of cognitive process is the sum of two additive processes is difficult to test, there is a great deal of evidence that the ex-Gaussian function provides a very good fit to several empirical RT distributions (Ratcliff & Murdock, 1976; Hockley, 1984; Luce, 1986). An interesting characteristic of the ex-Gaussian function is that its parameter values can easily be interpreted. Parameters μ and σ are the mean and standard deviation of the Gaussian component and can readily be interpreted as the central tendency and variability of RT for the decision making process. Parameter τ is the mean of the exponential component, which corresponds to the right 'tail' of the distribution; a larger τ implies a more skewed distribution. This has been interpreted as a measure of attention to understanding the problem prior to the neural circuitry attempting to solve it.

Simple mean response time tasks appear to have strong predictive power for performance on much more elaborate tasks, leading some theorists to propose that such RT measures grossly index the integrity or speed of processing in a way that benefits all tasks. An important line of work initiated by Schmiedek et al. (2007) shows that it is not just <u>average</u> RT that affects performance on a variety. Other statistics of the RT distribution are predictive of task performance and different statistics are predictive of performance that rely on different parts of cognition (e.g. processing speed, working memory, etc.). For example, Lövdén et al.. (2007) find that trial-to-trial variability in RT is correlated with cognitive decline in old age. Schmiedek et al. (2007) suggest that a the "worst performance rule" (related to τ_i), that is an individual's slowest reaction times across trials is predictive of short-term memory capacity, reasoning, and psychometric speed while average response time is related to the ability of a subject to hold attention. Gu et al. (2013) also finds that the parameter τ_i of the RT distribution estimated on the Conners' continuous performance test (CCPT), a test used to diagnose ADHD was sensitive to whether individuals were from a population with ADHD or were from a population of normal subjects. We note that Gu et al. do not investigate whether RT distribution parameters are predictive of future diagnosis of ADHD.

While the use of RT is common in psychology and neuroscience, economists in the past three years have become interested in such data. It is perhaps not a coincidence that to date use of RT data has been limited to experimental economics where data collection takes place in a laboratory (often on computers) as it does in psychology and neuroscience. For example, Krajbich et al. (2015) show that when subjects in a dictator game are given two divisions of tokens between self and other, it take subjects longest to make a choice when they value the two choices equally. What is of interest in this work is that because individuals differ in their social preferences, which choices are hard to make vary across subjects. For example, pro social subjects quickly choose to take 100 tokens for self and give 100 tokens to the other subject relative to taking 200 tokens for self and 0 tokens for the other subject. Selfish subjects also find this choice easy and quickly choose to take 200 tokens for self and give 0 tokens for the other subject. But a pro-social person struggles with (has long RT for) the choice between 80 for self and 100 for the other subject vs 100 for self and 80 for the other subject while a selfish person quickly chooses the later. A selfish person, on the other hand, has a long RT for the choice between 80 for self and 100 for the other subject vs 80 for self and 80 for the other subject; since selfish subjects place no weight on tokens to others they are indifferent between these two choices

Kranton et al. (2015) extend this analysis. They have subjects play 26 dictator game allocations twice, once where the other subject is a member of the dictator's political party

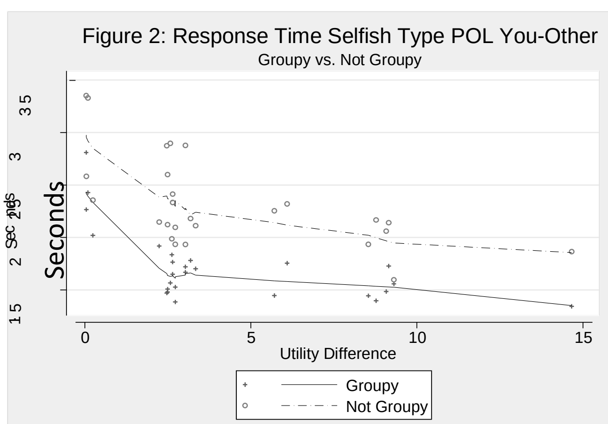


Figure 2: Response Time Selfish Type POL You-Other

(Democrat or Republican) and once where the other subject is a member of the opposing party. They randomize the order in which the party of the other subject appears. They show that some people are selfish regardless of the political party of the other subject ("not groupy"); but some people are prosocial when the other subject is a member of their political party but become selfish when the other subject is a member of the opposing party ("groupy"). Like Krajbich et al. (2015) they also find that RT is slowest when the difference in valuation between choices is small and becomes faster as the differences in valuation grow (Figure 2). In addition, they show (Figure 2) shows that groupy subjects, who must process the social context before making a choice take more time at any difference in valuation between choices even though they will end up acting selfishly. That is, having to judge the social context takes processing time above the time it takes to calculate the differences in valuation and to make a choice. This line of work suggests that RT data may be useful for a host of decisions including those of interest to economists. But to our knowledge this work has been restricted to the laboratory.

We use insights neuroscience as general guidance to our work but several caveats are in order. First, most cognitive neuroscience tasks are designed to have very fast responses, typically no more than 1 to 2 second. Second, most cognitive neuroscience tasks have forced binary choices. Finally, neuroscience tasks are designed to measure RT as the primary outcome and other processes that interfere with this measurement are carefully controlled in the laboratory. Surveys are quite different. The nondecision time can be quite extensive as questions can be long. Second, surveys are not designed to measure RT; RT is just recorded as a byproduct of collecting survey answers through a CAPI system. Finally, a host of issues and districtions in the field effect the speed at which respondents answer questions.

3. Data and Measurement of Cognition

This study uses data from the National Social, Health and Aging Project (NSHAP). NSHAP sampled 3,005 individuals in 2005/06 and is nationally representative study of U.S. residents ages 57 to 85. Wave 1 included a face-to-face interview (including a brief self-administered questionnaire), in-home collection of a broad panel of biomeasures, and a leave-behind questionnaire. Topics covered in the interview and leave-behind included self-reported health, physical function and morbidity, social networks, social support, marital history and intimate partnerships, sexuality, medication use, and demographic information. A second wave was collected in 2010/11 and a third in 2015/16. This study uses the 2,210 sample respondents that were surveyed in both wave 2 and 3 of NSHAP.

Importantly for our work, beginning in wave 2 NSHAP fielded an adapted version of the traditional paper and pencil assessment of the Montreal Cognitive Assessments (MoCA) that could be asked and answered on a survey using CAPI technology. The MoCA is a commonly used psychometric screening of cognitive functioning (Nasreddine et al., 2005). It assesses several different cognitive domains: attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. The MoCA-SA was designed for administration by non-medical personnel and to reduce respondent burden within the context of a large, time-limited national survey, while preserving the MoCA's sensitivity to a range of cognitive abilities. MoCA and MoCA-SA scores are highly correlated and scores can be accurately converted between the two scales (Kotwal, 2015). The specific items included are: 1) Orientation: date and month (2 points total); 2) Executive function: abstraction—similarity of watch and ruler (1 point), modified Trails-b (1 point); 3) Visuospatial skills: clock—contour, numbers, and hands (3 points total); 4) Memory: 5-word delayed recall (5 points); 5) Attention: forward digits (1 point), backward digits (1 point), subtract 7 s (3 points); and 6) Language: naming rhinoceros (1 point), phonemic fluency—words with the letter "F" (1 point for > 10 words in 60 s), and sentence repetition (1 point).

This validated, survey-adapted MoCA-SA, was fielded in the 2010/11 and the 2015/16 waves of NSHAP using the same CAPI system used for the entire interview. A great advantage of a CAPI system is that it records a time stamp when a respondent begins each question and ends each question. This allows analysts to know the RT on every survey question, including most items on the MoCA. Because of the nature of the MoCA-A, at times RT was collected for a set of tasks, such as the three sub-tasks of the clock draw (contour, numbers and hands). In total we have 14 separate response times for the 18 tasks. We view the overall MoCA-SA as a cognitive challenge, including each individual item, and we measure RT for individuals on it.

It is common for clinicians who use the MoCA for diagnosis to classify patients into three groups: Normal Cognition, Mild Cognitive Impairment (MCI) and Dementia. There is debate among clinicians as to what cutoff scores on the MoCA should be used to classify individuals into each group. Scores between 22 and 26 are typically used to distinguish Normal Cognition from MCI and scores between 17 and 20 to distinguish MCI from Dementia (which includes diagnoses including Alzheimer's disease (AD), Frontotemporal Dementia (FTD) and Vascular Dementia (VaD)).

It is useful to develop such cutoffs for the survey based MoCA-SA. There are many issues in doing this. Besides the uncertainty in cutoffs across studies, these cutoffs were constructed on clinical populations; many subjects were tested because cognitive issues had been detected. It is not clear that these same cutoffs would apply to a population representative sample. Secondly, for any MoCA cutoff we would adopt, this would need to be mapped to MoCA-SA scores. To do this, Kotwal, et al. (2015) collect the entire MoCA on a pilot sample of individuals and construct the standard MoCA score and the MoCA-SA score. Using this data they run a regression of the MoCA score against the MoCA-SA score. This regression suggests a mapping that MoCA = 1.14*MoCA-SA + 6.83. In this paper we use cutoffs of 25 to separate Normal Cognition from MCI and 19 to separate MCI from Dementia. Mapping this to MoCA-SA scores we classify respondent’s in NSHAP as has having Normal Cognitive functioning if they had a MoCA-SA score between 16 and 20, as having MCI if they had a MoCA-SA score between 12 and 15 and as having Dementia if they had a MoCA-SA score of 11 or lower.

Table 1 presents descriptive statistics on our sample. The top half of Table 1 presents the demographics of the sample while the bottom half present’s statistics on the MoCA-SA score in waves 2 and 3. The average age of our sample in wave 2 is nearly 72, with ages ranging from 62 to 90. More than half the sample is female and there is a wide dispersion of levels of education in the sample. Turning to the MoCA-SA score, we see the average score in wave 2 was 14.18; but

Table 1: Descriptive Statistics of the NSHAP Sample		
VARIABLES:	Mean	Std. Dev.
Demographics:		
Age	71.6	6.642
Fraction Female	0.548	0.498
Education		
Fraction Less than HS	0.174	0.379
Fraction HS Diploma	0.245	0.43
Fraction Some College	0.312	0.463
Fraction BA or above	0.269	0.443
MoCA SA Score:		
Wave 2		
MoCA SA	14.18	3.722
Fraction Normal	0.42	0.494
Fraction with MCI	0.359	0.48
Fraction with Dementia	0.22	0.415
Wave 3		
MoCA SA	13.36	4.119
Fraction Normal	0.355	0.479
Fraction with MCI	0.216	0.412
Fraction with Dementia	0.297	0.457

Notes: Tabulation based on 2,210 respondents who completed the MoCA-SA in waves 2 & 3 of NSHAP.

this obscures the fact that the most found levels of cognition remain in the normal range. 42% of subjects have normal cognitive functioning while 36% have scores that indicate MCI and 22% have scores that indicate dementia. With ageing between waves 2 and 3 cognition declines and the average MoCA-SA score in wave 3 is 13.36. What is notable is that the fraction with MCI falls from 36% to 22% while the fraction with Dementia rises from 22% to 30%

4. Results

We organize our results around three questions: (1) What is the relationship between response time on the MoCA-SA and performance on the cognitive test; (2) Are response times in wave 3 predictive of the level of performance on the MoCA-SA in wave 3 conditional on performance in wave 2; and (3) Does response time in wave 2 predict mortality between wave 2 and 3.

4.1 Response time and levels of cognition

Table 2 presents statistics on how long it took respondents to complete the 18 item MoCA-SA. On average respondents took 636 seconds (10 minutes 36 seconds to complete the MoCA-SA within the survey. There is a great deal of variation on how fast the MoCA-SA is completed with times to completion ranging from 1 minute 58 seconds to 60 minutes and one second. There is also a clear pattern of response time with respect to tested levels of cognition.

Subjects who had normal cognition had the fastest response times; they also had the lowest variance in response times; subjects who had MCI had slower response times and times with higher variance; and subjects who had dementia had the longest average response times and also the highest variance. While generally slower response times are correlated with lower levels of cognition, it is noteworthy that the minimum response time is highest for subjects in the normal cognition range and fastest for subjects in the dementia range. This is our first clue that in a survey context, response times are measuring a combination of processing speed to cognitive challenges and effort that the subject puts forth in meeting those challenges as well as other factors.

Table 2: Descriptive Statistics on Wave 2 Response Time to the MoCa-SA

Category	N	Mean	Std. Dev	Min	Max
Overall	2,209	636.2	196	118	3,601
Normal Cognition	928	594.4	154	269	2,126
MCI	794	644.8	208.6	174	3,601
Dementia	487	701.9	224.8	118	2,392

One factor that potentially is important is that NORC does not allocate survey takers to respondents randomly. Hard cases are typically handled by more experienced survey takers who are better able to carefully guide a respondent through the survey. It is possible that subjects with lower levels of cognition may appear as more difficult cases and hence survey takers who use more care and presumably more time may be assigned to respondents with lower levels of cognition. Column 1 of Table 3 repeats the mean levels of response time by level of cognition and reports the standard error while

Table 3: Wave 2 Response Time to the MoCa-SA with and without Interviewer Fixed Effects

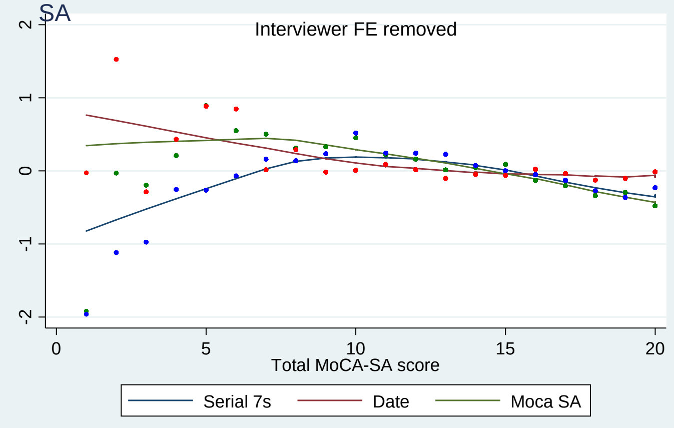
	Col (1)	**Col (2).**
Normal	594.4	595.3
	(6.3)	(5.3)
MCI	644.8	647.8
	(6.8)	(5.7)
Dementia	701.9	695.4
	(8.7)	(7.3)
Interviewer Fixed Effects	No	Yes

Note: T-tests of Dementia vs MCI, Dementia vs. Normal and MCI vs. Normal reject equality of means for both Col(1) and Col (2)

Column 2 removes interviewer fixed effects from the regression. Table 3 shows that interviewer fixed effects play a minor role in the differences in average response time across subjects with varying levels of cognition. What including interviewer fixed effects does appear to do however is to lower the variation of estimates. This suggests that while there are interviewers that are systematically faster and slower, their speed is largely uncorrelated with their assignment to subjects by levels of cognition.

Figure 3 displays the relationship between time to answer the MoCA-SA and the score on the MoCA-SA. In Figure 3, time to answer the MoCA-SA is displayed in standard deviation units. This is done so that the relationship between cognition and response time can be compared to two sub-components of the test – the serial 7s task and the date recall task. Figure 3 shows that time to complete the MoCA generally falls with the measured level of cognition. For example, subjects with scores of 20 complete the MoCA-SA about half a standard deviation faster than the sample average time while those with a score of 5 complete the MoCA-SA about half a standard deviation slower than the average. But Figure 3 also shows that this monotonic decline does not hold for all components. The monotonic decline holds strongly for asking respondents to recall the date (and for other easier tasks). But for the serial 7s task the monotonic relationship holds only for respondents that score 10 or above on the MoCA-SA. The serial 7s test takes a high degree of numeracy. Subjects that cannot subtract 7 or have sever memory issues simply give up on the task which produces very fast response times. These respondents tend to be those with very low MoCA-SA scores. For very difficult

Figure 3: Response Time (Z-Scores units) vs. W2 Moca SA



tasks, higher cognition within the low cognition range is associated with longer response times simply because higher cognition in this range allows some subjects to attempt the task.

As we discussed above, neuroscience models relate various moments of the response time distribution to levels of cognition. Table 4 presents a regression model of MoCA-SA against age, the time to complete the MoCA-SA in standard deviation units and a measure of the variability of the time to complete each item on the MoCA-SA. We classify a respondent as "steady" if on every item of the MoCA-SA the subject was above or below the average time in a consistent fashion. Specifically let r_{ki} be the time it takes for subject *i* to answer item *k* of the MoCA-SA. Let $s_{ki} = (r_{ki} - \overline{r_k})/\sigma_k$ where $\overline{r_k}$ is the mean time to complete task *k* and σ_k is the standard deviation of time to complete task *k*. That is s_{ki} is the time it takes *i* to answer item *k* in standard deviation units. The average amount of time in standard deviation units is then $\overline{s_i} = \sum_{k=1}^{14} s_{ki}/14$.[1] We then determine whether *i's* time to answer *k* was more than 1 standard deviation from *i's* overall mean. That is:

$$I_{ki} = \begin{cases} 1\ if -1 > (s_{ki} - \overline{s_i})\ or\ (s_{ki} - \overline{s_i}) > 1 \\ 0\ otherwise \end{cases} \tag{1}$$

We then sum over all items *k* for *i* and record

$$steady_i = \begin{cases} 1\ if\ \sum_{k=1}^{14} I_{ki} = 0 \\ 0\ otherwise \end{cases} \tag{2}$$

For example, if subject *i completed* the MoCA-SA in 12.6 minutes this was ½ a standard deviation above the mean time of 10.7 minutes; a subject was coded as "steady" if the subject was "about" ½ a standard deviations slower on each item of the MoCA-SA. Equations 1 and 2 implement the idea of "about" by scoring someone as steady on a specific item if the subject's time was within 1 standard deviation of the subjects own mean standardized time across all items.

[1] As explained above, while there are 18 items on the MoCA-SA, response times for some were measured for some tasks together, such as the three tasks associated with identifying aspects of a clock.

With these measures we implement the following regression model:

$$MoCA\_SA_i = \alpha + \beta_1 * Age_i + \beta_2 * Z_i + \beta_3 * Steady_i + \varepsilon_i \quad (3)$$

$$MoCA\_SA_i = \alpha + \beta_1 * Age_i + \beta_2 * Z_i + \beta_3 * Steady_i + \delta_k + \eta_i \quad (4)$$

where Age_i is *i's* age deviated from the sample mean age, Z_i is *i's* Z-score of wave 2 response time to the MoCA-SA, and δ_k are interviewer fixed effects. Table 4 displays several remarkable results from these regression models. Column 1 shows the estimated effect of Age and MoCA-SA response time measured in standard deviation units on the MoCA-SA wave 2 score. Column 1 shows the powerful correlation of the MoCA-SA response time on the MoCA-score. A one standard deviation increase in the amount of time to complete the MoCA-SA, for example increasing the time it takes to complete the MoCA-SA from the mean time 10.7 minutes to 14 minutes, is associate with a 0.645 point lower score. When we look at this effect relative to the effect of age, we see that a one standard deviation increase in the time it takes to complete the MoCA-SA has the same effect on the score as 5 years of age.

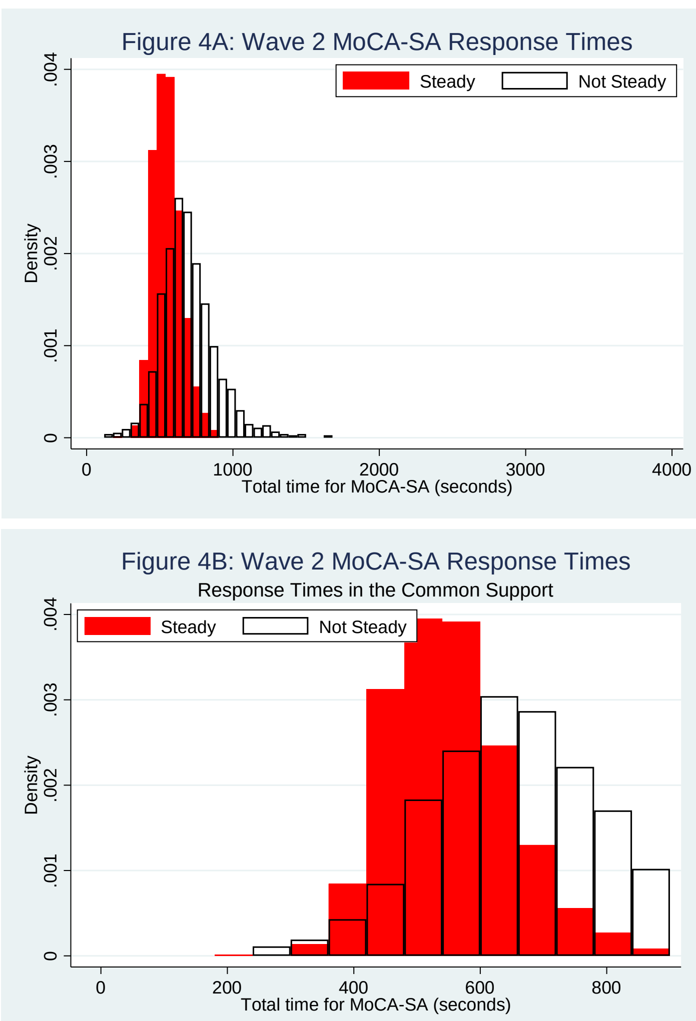
Table 4: Effects of Response Time Statistics on Wave 2 MoCa-SA Score with and without Interviewer Fixed Effects

VARIABLES	(1) MoCA-SA	(2) MoCA-SA	(3) MoCA-SA	(4) MoCA-SA
Age	-0.126***	-0.123***	-0.118***	-0.116***
	(0.0115)	(0.0114)	(0.0107)	(0.0106)
MoCA-SA time	-0.645***	-0.395***	-0.887***	-0.699***
	(0.0764)	(0.0816)	(0.0814)	(0.0870)
Steady		1.298***		0.920***
		(0.163)		(0.158)
Constant	14.18***	13.60***	14.18***	13.77***
	(0.0755)	(0.104)	(0.0677)	(0.0975)
Observations	2,209	2,209	2,209	2,209
R-squared	0.093	0.118	0.310	0.321
Interviewer FE	NO	NO	YES	YES

Notes: Age is deviated from the sample mean. MoCA-SA time is measured in standard deviation units.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In Column 2 we add our measure of the variability of response time. We see that answering the MoCA-SA items in a steady fashion is associated with substantially higher MoCA-SA scores. In general longer response times may also be associated with more variable response times across items and in fact we see that the effect of total response time is mitigated although it remains negative and strongly significant.

As we discussed before, it may be that more experienced interviewers are assigned to subjects with lower cognition. If this is the case then controlling for interviewer fixed effects could mitigate the effect of total MoCA-SA response time on the MoCA-SA score. Columns 3 and 4 repeat the models of Columns 1 and 2 but add interviewer fixed effects In fact, including interviewer fixed effects raises that effect of MoCA-SA response time on MoCA-SA score (Column 3). Interestingly, including interviewer fixed effects mitigates the estimated effect of answering MoCA-SA items in a steady fashion on the MoCA-SA score. This suggests that interviewers that are given cases where respondents have lower cognitive capacity are both faster and steadier in pacing the questionnaire than other interviewers.



One concern is that steady subjects likely are subjects that do not have extreme response times. Because of this, if entering response time linearly in the regression is not the right functional form, then steady might just be picking up non-linearity in the relationship between response time and cognition. Figure 4 A shows that steady individuals have lower response time on average; further there are response times for not steady individuals that lie outside the support of steady individuals. Figure 4B displays the response time for steady and not-steady subjects that are within the range of response times of steady subjects. At all times within this range, there are also response times for not steady subjects. But it is clear that even within the common support, not steady people are disproportionately at the upper end of the response time distribution. To balance the response time distribution we reweight the not steady subjects so that their response times will be distributed similarly to the steady subjects. The intuition of this is straight forward; looking at Figure 4B we see at 750 seconds about 20% of subjects are in the steady group and 80% in the not steady group. At 450 seconds about 80% of subjects are in the steady group and 20% are in the not steady group. If we were to down weight each of the not steady subjects at 750 seconds by 1/4th (20%/80%) and upweight the not steady subjects at 450 seconds by 4 (80%/20%) the weighted distribution of the not steady subjects would match the steady subjects. To implement this idea we estimate a propensity score logistic regression model where the dependent variable is whether you are steady or not steady and the independent variables are a cubic in response time. Having estimated the model we predict the probability of being in the steady group, p, at every response time in the common support. We then form a weight, w, where

$$w_i = \begin{cases} 1 \ if \ steady_i = 1 \\ p_i/(1-p_i) \ if \ steady_i = 0 \end{cases} \tag{5}$$

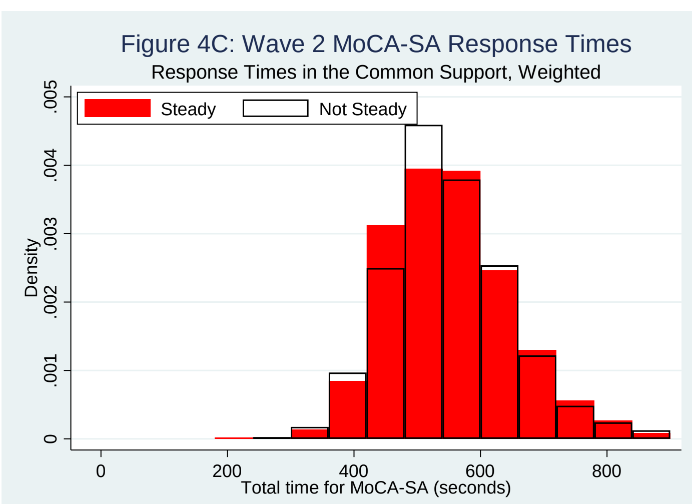


Figure 4C: Wave 2 MoCA-SA Response Times

Figure 4C shows the weighted distribution of response times on the common support. The propensity score model with a cubic in response time matches the distribution of response time across steady and not steady subjects well.

To check that our earlier regression results are not an artifact of the lack of common support between steady and not stead subjects, we re-estimate our models first using only subjects on the common support and then second, re-weighting the not steady subjects so that their response times match the response time distribution of the steady subjects. Table 5 shows these results. In general, focusing on the observations on the common support and reweighting observations for balance strengthens the estimated correlation of response time on MoCA-SA scores and weakens the correlation of being steady on MoCA-SA scores. For example, comparing Column 4 in Table 5 to Column 4 in Table 4 we see the effects of focusing on the common support and reweighting within a model with interviewer fixed effects. The correlation of MoCA-SA time with MoCA-SA score rises in absolute value from -0.699 to -1.382 while the correlation of steady falls from 0.920 to 0.369. While some quantitative difference arise when focusing on the common support, the qualitative result that both the average and variability of response time is strongly correlated with MoCA-SA score remains and results remain both substantively and statistically significant.

Table 5: Effects of Response Time Statistics on Wave 2 MoCa-SA, Respondents on the Steady Subject Support

	Unweighted		Propensity Weighted	
	(1)	(2)	(3)	(4)
VARIABLES	MoCA-SA	MoCA-SA	MoCA-SA	MoCA-SA
Age	-0.116***	-0.104***	-0.153***	-0.130***
	(0.0118)	(0.0110)	(0.0117)	(0.0110)
MoCA-SA time	-0.621***	-1.373***	-0.506***	-1.382***
	(0.134)	(0.149)	(0.154)	(0.171)
Steady	1.046***	0.563***	0.884***	0.369**
	(0.165)	(0.161)	(0.148)	(0.146)
Constant	13.76***	13.86***	13.95***	13.84***
	(0.106)	(0.0985)	(0.124)	(0.117)
Observations	2,033	2,033	2,033	2,033
R-squared	0.101	0.321	0.105	0.331
Interviewer FE	NO	YES	NO	YES

Notes: Age is deviated from the sample mean. MoCA-SA time is measured in standard deviation units.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.2 Response time and changes in levels of cognition

So far we have established that the total time and the variability in time across items of the MoCA-SA are strongly related to the level of cognition. This is consistent with findings in the neuroscience literature that it takes subjects less time to give correct than incorrect answers and that arriving at an answer is faster the lower the noise to signal. We now turn to a question that has been largely ignored in the neuroscience literature but is of great interest to the aging research community – can response times be used as a *predictor* or cognitive decline.

We start with a simple exercise. We break the sample up into 3 groups. Group 1 had no measured decline in MoCA-SA score in the five years between wave 2 and wave 3; group 2 had a mild decline of 1 to 2 points; and group 3 had a large decline of 3 or more points. We then calculate what the response time was on the wave 2 MoCA-SA for subjects in each group. Table 6 displays the average response times and the standard error for each average. What is clear is that subjects that experienced large declines in cognition between waves 2 and 3 took longer to answer the MoCA-SA in wave 2; that is long response times in wave 2 are a precursor to cognitive decline between waves 2 and 3. While the results are somewhat weaker when including interviewer fixed effects this general pattern holds.

Table 6: Wave 2 Response Time to the MoCa-SA, by the Decline in MoCA-SA Score between Waves 2 and 3, with and without Interviewer Fixed Effects

VARIABLES	N	Col (1)	Col (2)
None	1,026	612.2	624.5
		(5.6)	(4.6)
Mild (1 to 2)	631	642	633.1
		(7.3)	(6.4)
Large (3 or More)	552	674.2	661.6
		(9.7)	(8.3)

Note: All average times in column 1 are statistically significantly different; In column 2, average time for subjects with a large decline is statistically different than those with a mild or no decline; subjects with no and mild decline are not statistically different,

In order to judge how important wave 2 response times are relative to other factors such as age we run the following regressions:

$$MoCA\_SA\_W3_i = \alpha + \theta * MoCA\_SA\_W2_i + \beta_1 * Age_i + \beta_2 * Z_i + \beta_3 * Steady_i + \varepsilon_i \quad (5)$$

$$MoCA\_SA\_W3_i = \alpha + \theta * MoCA\_SA\_W2_i + \beta_1 * Age_i + \beta_2 * Z_i + \beta_3 * Steady_i + \delta_k + \eta_i \quad (6)$$

That is we ask how important is the standardized wave 2 response time in predicting wave 3 MoCA-SA <u>controlling for wave 2 MoCA-SA.</u> These results are displayed in Table 7. Column 1 includes only the standardized wave 2 response time and Column 2 adds our variable “steady.”

What is clear is that wave 2 response time is highly correlated with the MoCA-SA score in wave 3 conditional on the MoCA-SA score in wave 2.

Table 7: Wave 3 MoCA-SA as a function of Response Time to the MoCa-SA in Wave 2, with and without Interviewer Fixed Effects

VARIABLES	Col (1)	Col (2)	Col (3)	Col (4)	Col (5)	Col (6)
Age	-0.0927***	-0.0929***	-0.0947***	-0.0960***	-0.0959***	-0.0978***
	(0.00889)	(0.00887)	(0.00889)	(0.00913)	(0.00912)	(0.00914)
MoCA-SA time	-0.453***	-0.396***		-0.400***	-0.350***	
	(0.0584)	(0.0625)		(0.0694)	(0.0738)	
Steady		0.317**			0.263**	
		(0.126)			(0.133)	
MoCA-SA time *(Normal==1)			-0.171			-0.127
			(0.122)			(0.132)
MoCA-SA time *(MCI==1)			-0.435***			-0.388***
			(0.0936)			(0.1000)
MoCA-SA time *(Dementia==1)			-0.515***			-0.481***
			(0.107)			(0.119)
Steady *(Normal==1)			0.181			0.184
			(0.176)			(0.183)
Steady *(MCI==1)			0.549***			0.470***
			(0.170)			(0.177)
Steady *(Dementia==1)			0.326			0.129
			(0.268)			(0.278)
Observations	2,209	2,209	2,209	2,209	2,209	2,209
R-squared	0.581	0.582	0.584	0.612	0.612	0.614
Interviewer FE	NO	NO	NO	YES	YES	YES

Notes: Regressions control for MoCA-SA in wave 2
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The estimated impact of a 1 standard deviation increase in wave 2 response time is to lower wave 3 MoCA-SA score by over 0.4 points. This is equivalent to aging by more than 4 years or nearly doubling the cognitive aging process across the 5 years between waves. Column 2 shows that conditional on how long it took to answer the wave 2 MoCA-SA, being steadier across items was associated with a 0.32 point higher MoCA-SA in wave 3. Therefore, both the time it took to complete and the variability across items in the MoCA-SA in wave 2 are strongly predictive of the rate of cognitive decline between waves 2 and 3.

Column (3) estimates the correlation between our two statistics on wave 2 response time by the level of cognitive function in wave 2. We see that longer response times are especially correlated with cognitive decline when subjects have less than normal cognitive functioning in wave 2. It also appears that not being steady in answering the MoCA-SA in wave 2 is an especially important signal of looming cognitive decline for subjects displaying mild cognitive impairment in wave 2.

We now return to the issue raised earlier that some subjects that were not steady have response times that are outside the support of subjects that were steady. To investigate this we re-estimate equations 5 and 6 only on observations on the common support. We focus only on the model presented in Column 2 of Table 7 and the interviewer fixed effect equivalent in Column 5. The first two Columns of Table 8 present the results when only observations on the common support are used in the analysis. Columns 3 and 4 also apply the balancing weights. In general, the effect of MoCA-SA time is more strongly correlated with changes in MoCA-SA scores when we restrict our analysis to the common support. Now a one standard deviation in response time on the MoCA-SA has an equivalent effect on the wave 3 MoCA-SA as 10 years of aging. The evidence that not being steady is a precursor to cognitive decline is weaker in magnitude and in statistical significance.

Table 8: Effects of Response Time Statistics on Wave 3 MoCa-SA, Respondents on the Steady Subject Support

	Unweighted		Propensity Weighted	
	(1)	(2)	(3)	(4)
VARIABLES	MoCA-SA	MoCA-SA	MoCA-SA	MoCA-SA
Age	-0.0932***	-0.0960***	-0.0906***	-0.0929***
	(0.00931)	(0.00961)	(0.00944)	(0.00969)
MoCA-SA time	-0.702***	-0.782***	-0.854***	-0.938***
	(0.104)	(0.130)	(0.120)	(0.148)
Steady	0.218*	0.119	0.277**	0.141
	(0.128)	(0.138)	(0.116)	(0.124)
Observations	2,033	2,033	2,033	2,033
R-squared	0.568	0.600	0.556	0.601
Interviewer FE	NO	YES	NO	YES

Notes: Age is deviated from the sample mean. MoCA-SA time is measured in standard deviation units.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.3 Response time, health and mortality

A natural question is whether response time on cognitive tests can be used to model behavioral outcomes that are thought to be related to cognition.. Several studies suggest that cognitive impairment is a marker of physical decline and mortality, especially when dementia is present (Nguyen et al. (2003). Sacks (2009). There is some evidence that cognitive impairment is correlated with many dimensions of frailty in older adults including weight-loss, weakness, self-reported exhaustion, slow walking speed, and low physical activity level. While cognitive impairment may operate through frailty, Cano et al (2012) reports that cognitive impairment appears to have an independent effect on raising mortality risk. To the degree that our response time measures provide additional information on a respondent's cognitive ability and on its decline, it is of interest to know if these might provide independent markers of increased mortality risk.

In wave 3, NSHAP attempted to resurvey wave 2 respondents. When they were unable to resurvey individuals they asked a proxy respondent some questions including whether the original respondent was dead or alive and if alive whether they were in too poor health to participate in the survey. These three categories comprised the vast majority of cases. Of the wave 2 respondents, NSHAP successfully resurveyed 76%, 19% had died and about 5% had health conditions that prevented the respondent from participating in the survey.

We model these three outcomes with a multinomial logit model. Specifically, let we estimate

$$log(\pi_{2i}/\pi_{1i}) = \alpha_2 + \theta_2 * MoCA\_SA\_W2_i + \beta_{21} * Age_i + \beta_{22} * Z_i + \beta 2_{23} * Steady_i + \epsilon_i$$

$$log(\pi_{3i}/\pi_{1i}) = \alpha_3 + \theta_3 * MoCA\_SA\_W2_i + \beta_{31} * Age_i + \beta_{32} * Z_i + \beta 2_{33} * Steady_i + \epsilon_i$$

where π_{ki} is the probability of being in status $k \in \{1,2,3\} \equiv \{Alive, Dead, Poor\ Health\}$. Table 9 presents the base rates (Column 1) as well as the marginal effect for each variable of interest (Columns 2-5). First it is clear that the wave 2 level of cognition is highly correlated with the risk of death between waves. From a base rate of 18.9% a one standard deviation increase in the MoCA-SA score in wave 2 lowers the risk of death by 5.25 percentage points or by 28%. This is equivalent of more than 4 years of age. It also appears that wave 2 response time has an independent effect on mortality risk; a one standard deviation reduction in response time lowers the probability of death by 2.82 percentage points or by about 15%. Higher MoCA-SA scores and faster response times also lower the risk of being found in too poor health to take the wave 3 survey.

The results on whether the variance of response time effect mortality are weaker. For the models without interviewer fixed effects being steady lowers the probability of being alive at the 10% and the point estimates suggest it raises the probability of being dead or in poor health but these results are not statistically significant. Including interviewer fixed effects suggests that being steady raises the probability of death.

Table 9: Mortality and Poor Health as a function of Wave 2 MoCA-SA score and Response Time, with and without Interviewer Fixed Effects

	(1)	(2)	(3)	(4)	(5)	
VARIABLES	Base Rate	MoCA-SA (Z-score)	Age	Response Time (Z-score)	Steady	FE
Panel A: Without Interviewer Fixed Effects						
Alive	0.760***	0.0758***	-0.0155***	-0.0282***	-0.0209*	NO
	(0.00484)	(0.00477)	(0.000609)	(0.00523)	(0.0110)	
Dead	0.189***	-0.0525***	0.0129***	0.0195***	0.0171	NO
	(0.00460)	(0.00460)	(0.000607)	(0.00483)	(0.0105)	
Poor Health	0.0504***	-0.0233***	0.00267***	0.00871***	0.00379	NO
	(0.00270)	(0.00279)	(0.000371)	(0.00245)	(0.00630)	
Panel B: With Interviewer Fixed Effects						
Alive	0.760***	0.0896***	-0.0147***	-0.0267***	-0.0167	YES
	(0.00467)	(0.00522)	(0.000617)	(0.00624)	(0.0112)	
Dead	0.189***	-0.0601***	0.0122***	0.0186***	0.0242**	YES
	(0.00446)	(0.00506)	(0.000611)	(0.00587)	(0.0108)	
Poor Health	0.0504***	-0.0295***	0.00251***	0.00812***	-0.00753	YES
	(0.00259)	(0.00300)	(0.000379)	(0.00291)	(0.00641)	
Observations	6,270	6,270	6,270	6,270	6,270	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To investigate whether these results are sensitive to the lack of support in response times between steady and not steady subjects, we re-estimate our models for subjects on the common support, reweighting subjects so that the distribution of response times are balanced. These results are presented in Table 10. The effects of the MoCA-SA score remain unaffected and the effects of response time are larger. The effects of being steady on mortality risk now appear to be statistically insignificant.

Table 10: Mortality and Poor Health as a function of Wave 2 MoCA-SA score and Response Time, Observations on the Common Support Propensity Score Weighte, with and without Interviewer Fixed Effects

VARIABLES	(1) Base Rate	(2) MoCA-SA (Z-score)	(3) Age	(4) Response Time (Z-score)	(5) Steady	FE
Panel A: Without Interviewer Fixed Effects						
Alive	0.793***	0.0693***	-0.0144***	-0.0338***	-0.00933	NO
	(0.00507)	(0.00474)	(0.000641)	(0.00930)	(0.00973)	
Dead	0.168***	-0.0518***	0.0125***	0.0203**	0.00417	NO
	(0.00479)	(0.00450)	(0.000628)	(0.00872)	(0.00921)	
Poor Health	0.0387***	-0.0175***	0.00196***	0.0135***	0.00517	NO
	(0.00262)	(0.00254)	(0.000349)	(0.00455)	(0.00508)	
Panel B: With Interviewer Fixed Effects						
Alive	0.793***	0.0805***	-0.0139***	-0.0420***	-0.00517	YES
	(0.00485)	(0.00524)	(0.000654)	(0.0112)	(0.0101)	
Dead	0.168***	-0.0568***	0.0120***	0.0251**	0.00650	YES
	(0.00463)	(0.00500)	(0.000640)	(0.0105)	(0.00965)	
Poor Health	0.0387***	-0.0237***	0.00184***	0.0169***	-0.00133	YES
	(0.00247)	(0.00279)	(0.000367)	(0.00561)	(0.00519)	
Observations	5,850	5,850	5,850	5,850	5,850	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5. Conclusion

This paper examines whether response time data collected from CATI and CAPI systems is informative for modeling cognition, cognitive decline and outcomes related to cognition and cognitive decline. We use a unique opportunity as the NSHAP survey implemented a standardized cognitive assessment on a CAPI system. We investigate the total response time as well as a measure of item to item variability on the MoCA-SA fielded in wave 2 of the NSHAP. We find that the total response time is strongly related to measured cognitive level in wave 2, to the change in cognitive level between waves 2 and 3 and to mortality between waves. The variability in response times appears to be related to the level of cognition in wave 2 and in some specifications is related to the change in cognition between waves 2 and 3. The evidence that the variability in response times is related to mortality is weak.

This paper shows the usefulness of collecting item by item response times and using these to better understand respondent's capacity and behavior. The results have implications both for social scientists and clinicians. For clinicians, it is clear that useful information is being gathered when cognitive tests are administered that is not being used as only how well a patient did on the MoCA, or other cognitive tests, are used for diagnosis. But it is clear that how fast a patient got any score, and perhaps how variable the item-to-item responses were may be important for understanding cognitive aging. This suggests that combining information on response times and scores may help to better advice patients. Van der Linden (2007) offers a "Hierarchical framework for modeling speed and accuracy" simultaneously using Bayesian methods for estimation which seems a fruitful avenue to pursue for cognitive testing. We note that using both speed and accuracy to assess performance is common in the educational testing literature (see Lee and Chen (2011) for a review).

The implications for social scientists are also clear. Most survey collection organizations are quite familiar with surveys being a cognitive challenge. We have only looked at the amount of time and the variability across questions on a cognitive test to show in a survey context that response times can be useful for measuring cognition. But surveys are full of challenges including memory recall, numeracy and pattern recognition. It may be that response times to more traditional questions are also informative of cognitive ability. This would be exceptionally useful to know as Computer Assisted Telephone Interview (CATI) and CAPI systems have been capturing item by item response times since their use became widespread in the 1980s. This leaves open the exciting possibility of being able to use changes in response times to questions over time to index changing cognition of subjects. We note that it is rare that cognitive assessments are conducted in standard social science surveys, yet alone conducted in a repeated fashion. Therefore the potential to use response times collected as a by-product of normal survey operating procedures has the potential to give social scientists large representative samples with a repeated index of cognition. This could help enormously in increasing our understanding of how cognitive aging varies across the population.

Finally, this paper has just scratched the surface on the use of survey based response times. Just as neuroscientists gave us guidance on how these can be related to cognition, neuroscientists routinely use response times to measure other aspects of behavior. These included understanding valuation of choices, understanding deception and understanding non-cognitive skills such as effort. These are all areas that are of central importance to the social sciences.

References Cited

Baltes, P. B., & Mayer, K. U. (1999). *The Berlin Aging Study: Aging from 70 to 100*. New York, NY: Cambridge University Press.

Cano, C., Samper-Ternent, R., Snih, S. A., Markides, K., & Ottenbacher, K. J. (2011). Frailty and cognitive impairment as predictors of mortality in older Mexican Americans. The journal of nutrition, health & aging, 16(2), 142-147. doi:10.1007/s12603-011-0104-7

Couper, M.P., & Kreuter, F. (2013). „Using paradata to explore item level response times in surveys." *Journal of the Royal Statistical Society: Series A, 176*, 271-286.

Shoou-Lian Hwang Gu, Susan Shur-Fen Gau, Shyh-Weir Tzang, Wen-Yau Hsu, The ex-Gaussian distribution of reaction times in adolescents with attention-deficit/hyperactivity disorder, In Research in Developmental Disabilities, Volume 34, Issue 11, 2013, Pages 3709-3719, ISSN 0891-4222, https://doi.org/10.1016/j.ridd.2013.07.025.

Krajbich, I., et al. (2015). "Rethinking fast and slow based on a critique of reaction-time reverse inference." *Nature Communications* **6**: 7455.

Kranton et al. (2015) "Identity, Group Conflict, and Social Preferences." Duke University working paper.

Kotwal, A. A., Schumm, P., Kern, D. W., McClintock, M. K., Waite, L. J., Shega, J. W., ... & Dale, W. (2015). Evaluation of a brief survey instrument for assessing subtle differences in cognitive function among older adults. *Alzheimer Disease & Associated Disorders*, 29(4), 317-324.

Kotwal, A. A., Kim, J., Waite, L. J. & Dale, W. (2016). Social Function and Cognitive Status: Results from a US Nationally Representative Survey of Older Adults. *J Gen Intern Med,* DOI: 10.1007/s11606-016-3696-0

Langa, K. M., Plassman, B. L., Wallace, R. B., Herzog, A. R., Heeringa, S. G., Ofstedal, M. B., ... & Potter, G. G. (2005). The Aging, Demographics, and Memory Study: study design and methods. *Neuroepidemiology*, 25(4), 181-191.

Llewellyn, D. J., Lang, I. A., Langa, K. M., & Huppert, F. A. (2008). Cognitive function and psychological well-being: findings from a population-based cohort. *Age and Ageing*, 37(6), 685-689.

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.

Nasreddine, Z. S., Phillips, N. A., Dirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, 53(4), 695-699. doi:10.1111/j.1532-5415.2005.53221.x

Nathan, J., Wilkinson, D., Stammers, S., & Low, J. L. (2001). The role of tests of frontal executive function in the detection of mild dementia. *International Journal of Geriatric Psychiatry*, 16(1), 18-26.

Nilsson, L.-G., Adolfsson, R., Bäckman, L., de Frias, C. M., Molander, B., & Nyberg, L. (2004). Betula: A Prospective Cohort Study on Memory, Health and Aging. Aging, *Neuropsychology, and Cognition*, *11*, 134–148.

Nguyen, H. T., Black, S. A., Ray, L. A., Espino, D. V. and Markides, K. S. (2003), Cognitive Impairment and Mortality in Older Mexican Americans. *Journal of the American Geriatrics Society*, 51: 178–183. doi:10.1046/j.1532-5415.2003.51055.x

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive psychology*, 60(3), 127-157.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süss, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology. General*, *136*, 414–29.

Shega, J. W., Sunkara, P. D., Kotwal, A., Kern, D. W., Henning, S. L., Mcclintock, M. K., . . . Dale, W. (2014). Measuring Cognition: The Chicago Cognitive Function Measure in the National Social Life, Health and Aging Project, Wave 2. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 69(Suppl 2). doi:10.1093/geronb/gbu106

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, 44, 408–463.

Sachs, G. A. (2009). Dying from Dementia. *New England Journal of Medicine,* 361(16), 1595-1596. doi:10.1056/nejme0905988

Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *Journal of Clinical Psychiatry*.

Tse, C.-S., Balota, D. A., Yap, M. J., Duchek, J. M., & McCabe, D. P. (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology*, 24, 300–315.

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, *22*, 51-68.

Wagenmakers, EJ., Van Der Maas, H.L.J. & Grasman, R P.P.P. *Psychonomic Bulletin & Review* (2007) 14: 3. https://doi.org/10.3758/BF03194023.

Wallace, R. B., & Herzog, A. R. (1995). Overview of the Health Measures in the Health and Retirement Study. *The Journal of Human Resources*, 30. doi:10.2307/146279

Table 7: Wave 3 MoCA-SA as a function of Response Time to the MoCa-SA in Wave 2, with and without Interviewer Fixed Effects

VARIABLES	Col (1)	Col (2)	Col (3)	Col (4)	Col (5)	Col (6)
Age	-0.0927***	-0.0929***	-0.0947***	-0.0960***	-0.0959***	-0.0978***
	(0.00889)	(0.00887)	(0.00889)	(0.00913)	(0.00912)	(0.00914)
MoCA-SA time	-0.453***	-0.396***		-0.400***	-0.350***	
	(0.0584)	(0.0625)		(0.0694)	(0.0738)	
Steady		0.317**			0.263**	
		(0.126)			(0.133)	
MoCA-SA time *(Normal==1)			-0.171			-0.127
			(0.122)			(0.132)
MoCA-SA time *(MCI==1)			-0.435***			-0.388***
			(0.0936)			(0.1000)
MoCA-SA time *(Dementia==1)			-0.515***			-0.481***
			(0.107)			(0.119)
Steady *(Normal==1)			0.181			0.184
			(0.176)			(0.183)
Steady *(MCI==1)			0.549***			0.470***
			(0.170)			(0.177)
Steady *(Dementia==1)			0.326			0.129
			(0.268)			(0.278)
Observations	2,209	2,209	2,209	2,209	2,209	2,209
R-squared	0.581	0.582	0.584	0.612	0.612	0.614
Interviewer FE	NO	NO	NO	YES	YES	YES

Notes: Regressions control for MoCA-SA in wave 2

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1