

RESEARCH ARTICLE

Verbal fluency response times predict incident cognitive impairment

Matthew R. Ayers¹ | Justin Bushnell² | Sujuan Gao³ | Frederick Unverzagt⁴ |
John Del Gaizo⁵ | Virginia G. Wadley⁶ | Richard Kennedy⁶ | David Glenn Clark²

¹Department of Psychiatry, Richard L. Roudebush VA Medical Center, Indianapolis, Indiana, USA

²Department of Neurology, Indiana University, Indianapolis, Indiana, USA

³Department of Biostatistics, Indiana University, Indianapolis, Indiana, USA

⁴Department of Psychology, Indiana University, Indianapolis, Indiana, USA

⁵Biomedical Informatics Center, Medical University of South Carolina, Charleston, South Carolina, USA

⁶Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA

Correspondence

David G. Clark, MD, Department of Neurology, Indiana University, 355 W 16th Street, Suite 4020, Indianapolis, IN 46202, USA.
E-mail: clarkdg@iu.edu

Abstract

Introduction: In recent decades, researchers have defined novel methods for scoring verbal fluency tasks. In this work, we evaluate novel scores based on speed of word responses.

Methods: We transcribed verbal fluency recordings from 641 cases of incident cognitive impairment (ICI) and matched controls, all participants in a large national epidemiological study. Timing measurements of utterances were used to calculate a speed score for each recording. Traditional raw and speed scores were entered into Cox proportional hazards (CPH) regression models predicting time to ICI.

Results: Concordance of the CPH model with speed scores was 0.599, an improvement of 3.4% over a model with only raw scores and demographics. Scores with significant effects included animals raw and speed scores, and letter F speed score.

Discussion: Novel verbal fluency scores based on response times could enable use of remotely administered fluency tasks for early detection of cognitive decline.

KEYWORDS

Alzheimer's disease, cognitive impairment, cognitive neuropsychology, dementia, memory impairment, semantic memory, survival analysis, verbal fluency

Highlights:

- The current work evaluates prognostication with verbal fluency speed scores.
- These speed scores improve survival models predicting cognitive decline.
- Cases with progressive decline have some characteristics suggestive of Alzheimer's disease.
- The subset of acute decliners is probably pathologically heterogeneous.

1 | INTRODUCTION

The rising prevalence of dementia is an impending socioeconomic crisis.¹ One requirement for mitigating this crisis will be the development of inexpensive and widely accessible methods for identifying individuals at risk for cognitive decline. Biological markers of demen-

tia pathology become positive years before symptom onset and change dynamically with the course of disease,²⁻⁵ but tests for them are not accessible to all individuals.⁶ Cognitive tasks offer a means for reaching and identifying high-risk individuals, who may then undergo biomarker testing. Many cognitive tasks are freely available and easy to administer, even over the telephone, potentially overcoming limitations to

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* published by Wiley Periodicals, LLC on behalf of Alzheimer's Association

access. The elemental processes underpinning cognitive tasks are sensitive to early neuropathologic and biomarker changes, even in asymptomatic individuals.^{2,7-9}

In this study, we analyze recordings of telephone-administered verbal fluency tasks (VFTs) for identifying risk of incident cognitive impairment (ICI). In VFTs, patients are given 60 seconds to generate a list of words in response to a semantic cue (animals) or a letter cue (starts with "F"). The most commonly reported VFT score is the raw score, which consists of the number of unique, valid words generated in the allotted time. Emerging evidence suggests that VFT raw scores are useful across both preclinical and clinical stages of dementing illnesses. Asymptomatic, amyloid-positive subjects who convert to mild cognitive impairment (MCI) within 5 years exhibit faster rates of decline on follow-up testing with the animal fluency task than amyloid-negative subjects who remain cognitively and functionally normal.¹⁰ Baseline scores on animal and vegetable fluency tasks correlate with stability of MCI diagnosis 3 years later.¹¹ A recent meta-analysis suggests that baseline semantic fluency is among the most accurate predictors of incident conversion from MCI to dementia due to Alzheimer's disease (AD) and that it may be clinically useful for predicting a remote conversion to AD.¹² In subjects with AD, baseline impaired semantic performance, including lower scores on animal fluency, predicts increased rate of cognitive decline over the subsequent year.¹³

A broad range of cognitive factors may influence the raw score. These factors include semantic and lexical network integrity, strategic search and retrieval of appropriate exemplars, processing speed, and executive ability to monitor progress and remain in cognitive set. Scores indexing these cognitive processes may be a better reflection of specific, canonical brain-behavior relationships affected by various dementias.

One practical diagnostic approach has been to compare raw scores of different VFTs. Cognitively normal older adults maintain a semantic task advantage relative to letter fluency that persists into the ninth decade of life.¹⁴ Performance on the Mini-Mental State Examination (MMSE) and educational level do not account for this advantage¹⁴ and weakening of this advantage suggests the presence of AD rather than a vascular etiology of cognitive impairment¹⁵

Fewer studies have investigated the timings of VFT responses, but the idea is not new.¹⁶ A subset of studies analyzing VFT response timings focus on MCI or AD. This line of inquiry indicates that timing of animal fluency responses is helpful for discerning between MCI and AD,^{17,18} but most of these studies seek to resolve questions about cognitive processes underlying performance rather than questions about diagnosis or prognosis.¹⁹⁻²² A key result from previous work is that changes in timings may result from degradation of semantic networks or general mechanisms by which these networks are searched. Lenio et al.²¹ used a mathematical technique to separate effects from these two potential deficits and report evidence that either deficit may contribute to verbal fluency performance in MCI. These studies agree that in some cognitive disorders a systematic change occurs in verbal fluency performance as a function of task time and that features of this change likely reflect different neuropathological alterations.

RESEARCH IN CONTEXT

1. Systematic review: The authors reviewed the literature using traditional sources, including PubMed, Google Scholar, and Scopus. Publications reporting analyses of word retrieval times in verbal fluency tasks in healthy controls or in individuals with cognitive impairment, especially that deemed secondary to Alzheimer's disease, were examined. We appropriately cited and discussed articles we discovered through this review.
2. Interpretation: We discussed analyses of verbal fluency timing dating back as far as the 1940s, including those focusing on cognitive processes in healthy controls or conditions resulting in cognitive impairment. Our findings offer a valuable new approach to detecting the earliest changes of cognitive dysfunction using brief, remotely administered tests.
3. Future directions: Our next objective will be to examine these scores in individuals who have undergone more extensive clinical, neuropsychological, and biomarker evaluations.

Others have taken an approach to VFT analysis between the two outlined above, by dividing the allotted time into epochs and quantifying the number of valid words within each epoch. When analyzed in this way, VFT performance can distinguish between clinical AD and subcortical ischemic vascular disease (SIVD).²³ Further investigation along these lines shows that VFT epoch scores can discriminate cognitively normal controls, amnesic MCI, dysexecutive MCI, and mixed amnesic-dysexecutive MCI groups from one another.¹⁹

Inspired by these accounts, we assume that dementing diseases affect processes underlying verbal fluency performance and hypothesize that measurements of timing may help to detect and specify disease-related dysfunction in these processes. Moreover, given the statistically significant relationships between biomarker status and verbal fluency metrics in cognitively normal subjects we expect timing-based scores to discriminate subjects who will experience future cognitive decline.

Here, we explore a novel approach to the analysis of VFTs by capturing precise timings in the transcription of all spoken material in a relatively large cohort from the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study. This work builds on the timing approach taken in previous studies, but we take a view agnostic to the underlying cognitive mechanisms and focus instead on general patterns of VFT performance that may indicate early dysfunction in semantic or executive networks heralding overt decline. Our aim is to examine the intervals of silence between words produced and characterize how these timings, at a baseline evaluation, might signal impending cognitive decline. We use Cox proportional hazards (CPH) regression to investigate the hypothesis that scores derived from

inter-word intervals (IWI) add value to the raw score when predicting cognitive decline.

2 | SUBJECTS AND METHODS

2.1 | The REGARDS study

The REGARDS Study is a longitudinal epidemiological study that has enrolled and followed 30,239 individuals at least 45 years of age in the 48 contiguous United States.^{24–26} Enrollment is approximately evenly split between men and women and between White and Black races, with oversampling of the “stroke belt,” a region across the southeastern United States in which the incidence of stroke is increased relative to the rest of the country. REGARDS incorporates measures of cognition suitable for administration via telephone, including the Six-Item Screener (SIS)²⁷ and measures of verbal fluency. The SIS is a brief screening instrument consisting of three orientation questions and a three-item short-delay recall test. In a community-based sample, a cut-off score of < 5 points has a sensitivity of 74.2% for detection of cognitive impairment and a specificity of 80.2%.²⁷ Each REGARDS participant undergoes annual evaluation with the SIS. VFTs consist of giving the participant 60 seconds to generate a list of animals or words starting with the letter F. These tests are administered and recorded over the telephone biennially. The REGARDS Study has been approved by the institutional review board (IRB) at the University of Alabama at Birmingham. The data analysis was approved by the IRB of Indiana University.

2.2 | Subjects

We considered for inclusion REGARDS participants with no clinical stroke, who had undergone baseline assessment with animals and letter F VFTs, and defined cases of ICI based on three criteria applied to the vector of SIS scores: (1) all SIS scores up to the first evaluation after baseline VFT assessment must be normal, (2) at least one subsequent SIS score must be abnormal, and (3) the most recent SIS score must be abnormal. The first criterion ensures that the participant was cognitively normal at the time of the verbal fluency evaluation. The third criterion excludes individuals with an abnormal SIS score who had reverted to a normal score by the time of the data freeze. Controls were defined as having no abnormal SIS score at any evaluation. Based on these criteria, we were able to identify 641 case-control dyads, perfectly matched on the basis of age (difference ≤ 3 years), sex, four categories of educational level (less than high school, high school graduate, some college, and college graduate and above), race (White or Black), and three levels of geographic region (non-stroke belt, stroke belt, and stroke buckle). The stroke belt consists of 11 states across the southeastern United States, including Indiana, but excluding Florida, in which stroke mortality is higher than in other parts of the country.²⁸ The stroke buckle consists of a region west of the coastal plain of North Carolina, South Carolina, and Georgia, in which stroke mortality is higher

even than in other portions of the stroke belt.^{28,29} For cases, time to ICI was defined as the number of days from the baseline VFT evaluation to the first abnormal SIS score. For controls, time to ICI was censored at the time of the most recent SIS evaluation.

2.3 | Acute versus progressive decline

To examine verbal fluency patterns possibly arising from different pathological mechanisms, we divided the ICI participants into those who exhibited a progressive decline and those who declined acutely. Importantly, these definitions depended on examination of each individual's entire array of SIS scores—eight to nine annual SIS evaluations on average, some of which preceded the verbal fluency evaluations. We defined a “downward step” as any transition to an abnormal score that is also a new minimum score for that participant. We then defined progressive decline to be the presence of > 1 downward step to reach the lowest SIS score and acute decline as a transition from any normal SIS score to the lowest recorded SIS score in a single downward step. Note that in all progressive cases and some acute cases, the designation of the pattern of decline depended on observations occurring after the first abnormal SIS score, that is, after the transition to ICI was marked for the purpose of the survival analysis.

2.4 | Data pre-processing

We transcribed 2364 verbal fluency recordings by first obtaining a rough transcription with the Amazon Web Services (AWS) Transcribe tool and then correcting the initial transcriptions by means of a MATLAB-based graphical transcription tool developed in our lab for this purpose. With this tool, we were able to amend the transcriptions with an onset time and duration for each utterance, and from these timings the calculation of IWI was straightforward.

2.5 | Statistical analyses

2.5.1 | Predictors of time to ICI

We examined predictors of time to ICI conversion (in days) by means of CPH regression. Predictors included traditional raw scores (the number of valid items generated by each participant in 60 seconds) and speed scores for each VFT based on IWI. The IWI measures themselves were severely positively skewed and were normalized by taking the fourth root. To capture speed (as opposed to the response latency reflected by IWI, which is actually a measure of slowing), we inverted the fourth root-transformed IWI in two steps. First, we applied min-max normalization to the entire vector of IWI for each task. Minmax normalization of a vector entails subtracting the minimum value from every entry and then dividing the entries by the difference between the minimum and the maximum. This transformation places all of the vector elements on the interval from 0.0 to 1.0. Second, we subtracted

TABLE 1 REGARDS participant data

	Controls (N = 641)	ICI (N = 641)	Acute decliners (N = 537)	Progressive decliners (N = 104)
Age (years)	74.94 (8.62)	74.96 (8.72)	74.40 (5.23)	77.87 (8.10) ^{*,a}
Sex (F:M)	347:294	347:294	277:260	70:34 ^{*,a}
Region (NB:BELT:BUCKLE)	258:243:140	258:243:140	219:201:117	39:42:23
Education (< HS:HS:SC:CG+)	75:191:171:204	75:191:171:204	60:167:141:169	15:24:30:35
Race (B:W)	244:397	244:397	209:328	35:69
Time to conversion or censoring (days)	1081.32 (451.35)	1000.69 (468.36) [*]	1022.93 (481.60) [*]	885.88 (374.35) ^{*,a}
Number of SIS assessments	8.85 (2.03)	8.81 (2.04)	8.73 (2.11)	9.19 (1.59) ^a
Minimum SIS score	5.17 (0.38)	3.27 (1.12) [*]	3.54 (0.90) [*]	1.86 (1.07) ^{*,a}
Framingham stroke risk score	10.94 (9.61)	11.66 (9.59)	11.31 (9.67)	13.48 (9.02) [*]

Abbreviations: B, Black; CG+, college graduate or above; F, female; ICI, incident cognitive impairment; M, male; NB, nonbelt; HS, high school; REGARDS, Reasons for Geographic and Racial Differences in Stroke Study; SC, some college; SIS, Six-Item Screener; nW, White.

Notes: Metric variables are shown as mean (standard deviation) and compared to *t*-test. Categorical variables were compared using χ^2 .

^{*}*P* < .05 compared to controls with *t*-test or χ^2 ;

^a*P* < .05 compared to acute decliners. For Framingham stroke risk, controls, ICI, acute, progressive groups have *n* = 574, 552, 463, and 89, respectively.

the minmax normalized IWI from 1.0, so the fastest transition in the set would receive a score of 1.0 and the slowest transition a score of 0.0. Adding these speed measurements for each individual word list, we derived a speed score, which one may view as a form of the raw score that was transformed to penalize lengthy pauses during the task. Because this score was strongly correlated with raw score (> 0.9) we avoided multicollinearity effects by regressing each speed score onto both raw scores (animals and letter F) and using the residual for the speed score when fitting models. We performed comparisons between the various groups of participants using traditional *t*-tests and Bayes factors.³⁰

We began by fitting a base model with only VFT raw scores (animals and letter F) and demographic variables. The decision to include both raw scores was based on the expectation that including two similar scores would improve the overall reliability and statistical stability of the models. We then fit a second model by adding speed scores to the base model and compared the models in terms of concordance, that is, the probability that the model, given a randomly selected case and control, will assign a higher risk estimate to the case. We used a *z*-test to evaluate the statistical significance of improvement in concordance. To examine patterns among the acute and progressive subsets of cases, we re-fit the best model twice. Both of the additional models included all controls, but the first included only the acute decliners and the second only the progressive decliners.

3 | RESULTS

Table 1 shows data on the 641 clinically stroke-free individuals with ICI and matched controls. (See Figure S1 in supporting information for a flowchart depicting participant selection.) We observed the progressive pattern in 104 cases and the acute pattern in 537 cases. There were no significant differences between cases and controls in terms of

age, sex, education, region, or race. Time to ICI conversion was shorter than time to censoring for controls. Minimum SIS score was, of necessity, lower in the cases. Cases had a slightly higher Framingham stroke risk (FSR) score than controls, but the difference was not statistically significant.

Separation of cases into acute and progressive decliners yielded the same pattern of results when the acute group was compared to the controls. However, progressive decliners were older than both controls and acute decliners, and more likely to be female. The progressive group converted to ICI more rapidly than the acute group (861.70 days vs. 1015.41 days). Individuals in the progressive group underwent more SIS assessments than those in the acute group and reached a lower overall SIS score (1.84 in progressive vs. 3.54 in acute). The progressive group had a higher FSR score than controls.

Table 2 shows the VFT raw and speed scores for the groups and results of traditional frequentist statistical comparisons and Bayes factor comparisons. Because of the large variation in Bayes factors, we report the log-transformed Bayes factors (log-BF), such that a value of zero indicates an absence of evidence supporting either the null or alternative hypotheses, a log-BF < -1 indicates strong evidence in favor of the null (i.e., no difference between groups), and a log-BF > 1 indicates strong evidence in favor of the alternative. ICI cases performed worse than controls on all four scores, with log-BF values uniformly greater than 1. On the whole, the subset of acute decliners exhibited a pattern similar to that of the ICI group as a whole. However, when this group was compared to controls, the log-BF values were smaller for three of the four scores and larger for the letter F raw score (1.40 vs. 1.15). With regard to the two animal fluency scores, the subset of progressive decliners exhibited log-BF values very similar to those of the acute group, but differed starkly in terms of the letter F scores. For letter F raw score, the log-BF indicated moderate evidence in favor of the null hypothesis (log-BF = -0.74), and only weak evidence in favor of a lower letter F speed score than controls (log-BF = 0.18). Figure 1

TABLE 2 Comparison of verbal fluency scores among groups

	Controls	ICI	Acute	Progressive	log BF compared to controls			log BF progressive versus acute
					ICI	Acute	Progressive	
Animals raw score	16.62 (5.33)	15.13 (5.15)*	15.29 (5.23) [†]	14.31 (4.62)*	4.34	2.79	2.69	-0.27
Animals speed score	8.54 (3.56)	7.26 (3.29)*	7.41 (3.37) [†]	6.49 (2.72)* ^a	9.19	6.34	6.54	0.50
Letter F raw score	11.63 (4.51)	10.78 (4.69)*	10.70 (4.63) [†]	11.17 (4.97)	1.15	1.40	-0.74	-0.74
Letter F speed score	4.89 (2.57)	4.25 (2.55)*	4.25 (2.59) [†]	4.26 (2.51)*	3.08	2.73	0.18	-0.92

Abbreviations: BF, Bayes factor (presented as the base-10 logarithm due to extreme variation); ICI, incident cognitive impairment.

* $P < .05$; ^a indicates $P < .05$ compared to acute decliners.

Note: log BF > 0 suggests at least weak evidence for effect, > 1 suggests strong evidence for effect, < 0 suggests at least weak evidence against the presence of an effect, < -1 suggests strong evidence against the presence of an effect.

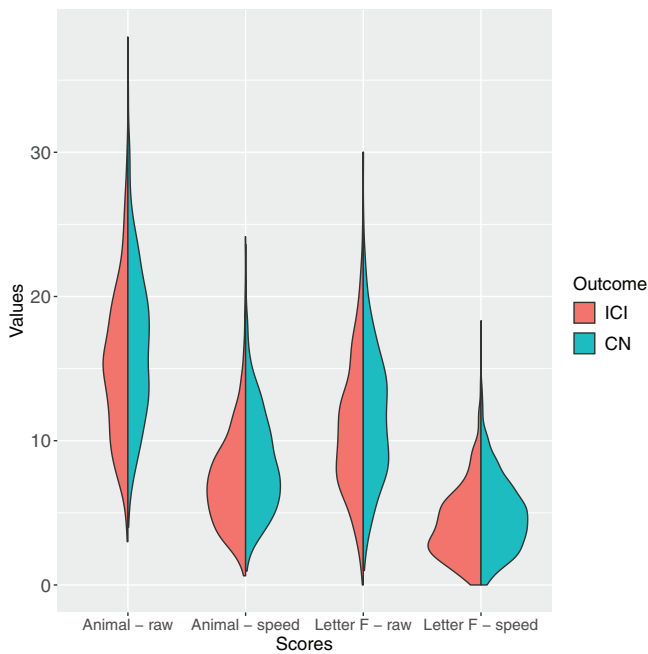


FIGURE 1 Split violin plots showing distribution of animals and letter F raw and speed scores. Each side of a “violin” consists of the probability density function for one of the groups, smoothed using kernel density estimation. ICI, incident cognitive impairment (case); CN, cognitively normal (control)

depicts split-violin plots comparing scores on each VFT measure for ICI cases and controls.

Comparison of the acute and progressive decliner subsets revealed no significant effect and suggested at least weak evidence in favor of the null hypothesis for three of the four scores (log-BF < 0), with the exception being animal fluency speed score, for which the t-test was significant ($t[107.86] = -3.03, P < .01$) and log-BF indicated moderate evidence (0.50). Figure S2 in supporting information depicts split-violin plots comparing VFT scores between the acute and progressive decliner subsets.

The base model, which included only raw scores and demographic variables, had a concordance of 0.565 (see Table 2). Higher values of animals raw score were associated with lower risk of ICI (hazard ratio [HR] = 0.96 with 95% confidence interval [0.94–0.98], $Z = -4.53,$

$P < .0001$). None of the demographic variables made a significant contribution to the model (Table 3).

Table 4 shows the model resulting from addition of letter F and animals speed scores to the base model. Addition of these scores led to a 3.4% improvement in concordance to 0.599 ($Z = 3.11, P < 0.002$). Animals raw score and speed score made independent, significant contributions (raw: HR = 0.96 [0.94–0.97], $Z = -5.05, P < 0.0001$; speed: HR = 0.85 [0.80–0.90], $Z = -5.35, P < .0001$). Letter F speed contributed significantly (HR = 0.92 [0.85–0.99], $Z = -2.09, P = .04$).

Separation of the acute and progressive decliners revealed a few qualitative differences in terms of predictors of risk (Table 5). While animals raw and speed scores remained strong predictors for both groups (all P -values < .0001), letter F speed score was predictive of decline only in the progressive group (HR = 0.80 [0.67–0.97], $Z = -2.33, P = .02$). Age and sex were significantly associated with risk only in the progressive group (age: HR = 1.04 [1.01–1.07], $Z = 2.81, P = .005$; sex: HR = 0.58 [0.38–0.89], $Z = -2.50, P = .01$).

4 | DISCUSSION

Successfully confronting the imminent increase in dementia prevalence will require screening methods with a number of desirable qualities, including wide accessibility with minimal invasiveness, burden, and cost. Brief, telephone-administered neuropsychological tests may provide an inexpensive approach to remotely screening large numbers of at-risk individuals for dementia research purposes,³¹ as well as accommodating the current emphasis on telehealthcare. Several studies suggest that verbal fluency raw scores are valuable for detecting early cognitive changes of AD.^{15,32–34} An efficient method for improving on verbal fluency raw scores for the purpose of detecting and differentiating dementia in its earliest stages would be a step forward for the field.

In this nested case-control study, we aimed to characterize the prognostic value of telephone-based VFT with a novel approach accounting for performance as a function of time. We transcribed audio recordings of REGARDS participants performing VFTs at baseline, amending the transcriptions with timing measurements of each utterance. After fitting a base model with only raw scores and demographic variables, we added speed scores and evaluated for increase in concordance. We then evaluated models stratified by acute and progressive decline.

TABLE 3 Cox proportional hazards—Raw verbal fluency scores (N = 1282)

Variable	Coefficient	SE	Hazard ratio	CI	Z	P-value
Animals—raw	−0.04	0.009	0.96	(0.94–0.98)	−4.53	< 0.0001
Letter F—raw	0.00	0.01	1.00	(0.98–1.02)	0.08	0.94
Age	0.01	0.005	1.01	(1–1.02)	1.36	0.17
Sex (M)	−0.03	0.082	0.97	(0.83–1.15)	−0.31	0.76
Some college	0.15	0.145	1.17	(0.88–1.55)	1.07	0.29
College+	0.08	0.148	1.08	(0.81–1.44)	0.53	0.60
HS graduate	0.09	0.14	1.10	(0.83–1.44)	0.66	0.51
Race (W)	0.02	0.09	1.02	(0.85–1.22)	0.22	0.83
Region (BELT)	−0.13	0.093	0.88	(0.73–1.05)	−1.42	0.16
Region (BUCKLE)	−0.04	0.107	0.97	(0.78–1.19)	−0.33	0.75

Abbreviations: CI, 95% confidence interval of hazard ratio; HS, high school; M, male; SE, standard error; W, White.
Note: Concordance = 0.565.

TABLE 4 Cox proportional hazards—Speed scores added to model

Variable	Coefficient	SE	Hazard ratio	CI	Z	P-value
Animals—raw	−0.04	0.01	0.96	(0.94–0.97)	−5.05	< 0.0001
Letter F—raw	0.00	0.01	1.00	(0.98–1.02)	0.09	0.93
Animals—speed	−0.17	0.03	0.85	(0.80–0.90)	−5.35	< 0.0001
Letter F—speed	−0.08	0.04	0.92	(0.85–0.99)	−2.09	0.04
Age	0.01	0.01	1.01	(1–1.02)	1.33	0.18
Sex (M)	−0.04	0.08	0.96	(0.82–1.13)	−0.48	0.63
Some college	0.17	0.14	1.18	(0.89–1.57)	1.14	0.25
College+	0.09	0.15	1.10	(0.82–1.46)	0.63	0.53
HS graduate	0.07	0.14	1.07	(0.81–1.41)	0.50	0.62
Race (W)	−0.01	0.09	0.99	(0.83–1.18)	−0.13	0.90
Region (BELT)	−0.11	0.09	0.89	(0.75–1.07)	−1.21	0.23
Region (BUCKLE)	−0.06	0.11	0.95	(0.77–1.17)	−0.52	0.61

Abbreviations: CI, 95% confidence interval of hazard ratio; HS, high school; M, male; SE, standard error; W, White.
Note: Concordance = 0.599.

Addition of speed scores improved on the base model for detecting risk for cognitive decline, with an increase in concordance of 3.4%. We achieved the highest concordance (0.743) by fitting a model on the progressive subgroup separately. The improvement in concordance is small, but statistically robust, and raises the possibility that timing-based features could make an important contribution to more comprehensive machine learning models trained to stratify individuals in terms of dementia risk based on VFT. Our findings are consistent with previous research on the temporal characteristics of connected speech in MCI and AD.^{35–37}

From the standpoint of disease prevalence, AD is likely the most frequent pathology underlying transitions to ICI in our sample. According to pathological studies of AD progression,³⁸ the microtubule associated protein tau begins accumulating in the entorhinal cortex and spreads readily into the hippocampus. Early involvement of the perirhinal cortex likely contributes to early semantic dysfunction in AD.^{9,39} Subsequent Braak stages are marked by spread of tau into neocorti-

cal regions, including the inferior temporal lobe, and middle and superior temporal gyri.³⁸ These regions are also important for normal language function, including semantic memory of animal concepts and words.^{40,41} We propose that pathologic changes in these neural networks impact dynamics of word retrieval in ways that are readily detected in the timings of retrieved words during verbal fluency. The assumption that AD is prevalent in our sample fits well with the view that many of our results could arise from lexical-semantic disruption. Based on the theoretical organization of lexical networks we expected the lexical characteristics of adjacent words to influence IWI duration and IWI-related variables. This conjecture finds support in previous work comparing the influence of these forms of similarity on verbal fluency.⁴²

In keeping with the view that many of our cases are suffering from AD and AD impacts lexical-semantic networks, we observed that animal fluency raw and speed scores contributed significantly to estimates of risk in all of our CPH models, regardless of which other predictor

TABLE 5 Cox proportional hazards analysis—Stratified by acute versus progressive decline

Progressive decline						
Variable	Coefficient	SE	Hazard ratio	CI	Z	P-value
Animals—raw	−0.11	0.02	0.90	(0.86–0.94)	−4.47	< 0.0001
Letter F—raw	0.01	0.03	1.01	(0.96–1.06)	0.44	0.66
Animals—speed	−0.37	0.08	0.69	(0.59–0.8)	−4.68	< 0.0001
Letter F—speed	−0.22	0.09	0.80	(0.67–0.97)	−2.33	0.02
Age	0.04	0.01	1.04	(1.01–1.07)	2.81	0.005
Sex (M)	−0.54	0.22	0.58	(0.38–0.89)	−2.50	0.01
Some college	0.16	0.33	1.17	(0.62–2.24)	0.49	0.62
College+	0.03	0.34	1.03	(0.53–1.99)	0.08	0.4
HS graduate	−0.40	0.34	0.67	(0.34–1.31)	−1.17	0.24
Race (W)	0.28	0.23	1.33	(0.85–2.06)	1.25	0.21
Region (BELT)	0.18	0.23	1.20	(0.77–1.88)	0.80	0.42
Region (BUCKLE)	0.06	0.27	1.06	(0.63–1.79)	0.21	0.83
Acute decline						
Animals—raw	−0.04	0.01	0.96	(0.94–0.98)	−4.15	< 0.0001
Letter F—raw	0.00	0.01	1.00	(0.98–1.02)	−0.34	0.74
Animals - speed	−0.16	0.03	0.85	(0.8–0.91)	−4.71	< 0.0001
Letter F—speed	−0.07	0.05	0.93	(0.85–1.02)	−1.56	0.12
Age	0.00	0.01	1.00	(0.99–1.01)	0.53	0.60
Sex (M)	0.04	0.09	1.04	(0.87–1.24)	0.46	0.64
Some college	0.21	0.16	1.23	(0.9–1.69)	1.30	0.19
College+	0.13	0.16	1.13	(0.82–1.56)	0.77	0.44
HS graduate	0.16	0.16	1.17	(0.87–1.59)	1.03	0.30
Race (W)	−0.04	0.10	0.97	(0.8–1.17)	−0.36	0.72
Region (BELT)	−0.16	0.10	0.86	(0.7–1.04)	−1.53	0.13
Region (BUCKLE)	−0.07	0.12	0.93	(0.74–1.17)	−0.60	0.55

Abbreviations: CI, 95% confidence interval of hazard ratio; HS, high school; M, male; SE, standard error; W, White.
Notes: Concordance of progressive model = 0.743. Concordance of acute model = 0.593.

variables were included. The finding that animal speed scores improve predictions of ICI extends the finding of Koenig et al., who observed that durations of silence during the first 15 seconds of animal fluency aided in the discrimination of MCI and AD, but not of MCI and healthy controls.¹⁷ We find further support for the prognostic value of VFTs in that letter F speed scores are negatively associated with ICI risk, a finding that appears to be driven by the subset with progressive decline. This effect is also likely to be due to degeneration of lexical-semantic networks, as semantic effects have been observed to play an important role in ordering of words during letter fluency.⁴²

The approach described could aid the development of models that differentiate among underlying causes of cognitive impairment. The progressive decliners were older and more likely to be female than the acute decliners (Table 1). Both of these demographic findings have been described among AD cohorts.⁴³ The finding of higher FSR score in this group (compared to controls) is equivocal. On one hand, stroke risk factors are associated with increased risk for AD,⁴⁴ but on the other hand, age is a component of the FSR, and the differences in age are there-

fore contributing to differences in FSR. In addition, by design, FSR is associated with increased risk for stroke, and stroke contributes to risk for both AD and non-AD dementia.^{43,45} Finally, letter F speed score proved to be a significant predictor of progressive decline, but not acute decline, possibly indicating a different pattern of lexical-semantic disturbance.

The evidence suggests that the acute decliners are likely to be a pathologically heterogeneous group. We originally thought that sub-clinical stroke would be a major contributor to acute decline, but the fact that the acute group did not differ from controls on FSR argues against this supposition. In addition, based on prevalence data, we would not expect cases of vascular cognitive impairment to outnumber cases of cognitive impairment due to AD by a 5:1 ratio. Along these lines, we would expect perhaps 60% of the ICI cases to be due to AD, and the fact that only 16% fell into the progressive group suggests that some members of the acute group are likely to have AD. Comparing results from CPH models, there is no qualitative difference between acute and progressive groups in animal fluency speed

and raw score contributions, suggesting that delays in access to animal names are comparable between the two groups. There are two potential explanations for this observation. One possibility is that the cascade described above, of tau spreading into neocortical regions supporting animal word and concept knowledge, is also occurring among at least some members of the acute group. If so, then these members of the acute group would likely be positioned at an earlier time point (than those in the progressive group) on the trajectory toward ICI. This view is supported by the finding that members of the progressive group were older than the acute group and converted an average of 6 months earlier. The second possibility is that some other mechanism is playing a role in generating this pattern. Candidate mechanisms would include formulation and execution of lexical search strategies dependent on frontal-subcortical circuits, as proposed by Lenio et al.²¹ Executive dysfunction does contribute to cognitive decline in all three of the most prevalent causes of dementia (AD, Lewy body dementia [LBD], and cerebrovascular disease [CVD]), although it tends to be more severe in LBD and CVD.^{46–49} Performance on the letter fluency task is generally considered to rely more heavily on executive function than animal fluency.¹⁵ We suspect that the significant contribution of letter F speed score to progressive ICI reflects changes occurring later on the decline trajectory, but cannot exclude the possibility that it represents a clue to underlying pathology.

The main limitations in this work are the lack of neurodegenerative biomarkers and lack of a complete neurocognitive assessment. If we had had access to additional data of these kinds, it might have been possible identify a more pathologically homogeneous cohort. By design, the SIS is nonspecific for etiology and correlates with global cognitive impairment.²⁷ The SIS may not be a difficult enough task for MCI subjects and may therefore produce a ceiling effect.⁵⁰ Insensitivity of the SIS could contribute to heterogeneity in the sample of cognitively impaired individuals. This heterogeneity is likely to have been compounded by the lack of biomarkers in our analysis. We have begun to address this limitation by partitioning the ICI group based on clinical course, but future research will assess the effects we report here in research participants with biomarker data and detailed cognitive assessment. A second limitation is that the speed scores we have characterized may be difficult to generalize to other samples. Application of the method requires identification of valid words and measurement of the intervals between them. In our case, this work was painstaking, but advances in automatic speech recognition could render it trivial. Other outstanding questions are whether the mathematical transformations we used may be applied unmodified to new data sets. These transformations included first the normalization of the IWI by taking the fourth-root, and second the minmax normalization procedure. As a final note, the speed scores as we defined them do not take into account phenomena such as “burstiness”⁵¹ or “clustering,”⁵² but these phenomena will be addressed in future work.

5 | CONCLUSION

We have presented a novel approach to the analysis of verbal fluency recordings. The key finding is that addition of scores derived from word

timings offer an improvement over traditional fluency raw scores. In addition, we find that time to cognitive decline is negatively associated with the letter and animal fluency speed score, as well as animal fluency raw score. This pattern is present across the sample as a whole, but separation of cases into acute and progressive decliners reveals that letter F speed score has greater value for predicting progressive decline. These discrepancies raise the question of whether the acutely declining subset are pathologically distinct or are simply being tested at an earlier point on the trajectory toward cognitive impairment. Further study is warranted to clarify this question, as either possibility would be important for the application of verbal fluency recording analysis to the problem of early dementia detection.

FUNDING SOURCES

NIH U01 NS041588.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

REFERENCES

1. Brookmeyer R, Gray S, Kawas C. Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *Am J Public Health*. 1998;88(9):1337-1342.
2. Bateman RJ, Xiong C, Benzinger TLS, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med*. 2012;367(9):795-804.
3. Jack CR Jr, Bennett D, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's Dement*. 2018;14:535-562.
4. Mormino EC, Papp KV, Rentz DM, et al. Early and late change on the preclinical Alzheimer's cognitive composite in clinically normal older individuals with elevated amyloid beta. *Alzheimer's Dement*. 2017;13(9):1004-1012.
5. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement : J Alzheimer's Assoc*. 2011;7(3):280-292.
6. Laske C, Sohrabi HR, Frost SM, et al. Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimer's Dement*. 2014: 1-18.
7. Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology*. 2013;80:1778-1783.
8. Diamond A. Executive functions. *Annu Rev Psychol*. 2013;64:135-168.
9. Ferretti MT, Iulita MF, Cavedo E, et al. Sex difference in Alzheimer's disease- the gateway to precision medicine. *Nat Rev Neurol*. 2018;14:457-469.
10. Howieson DB, Carlson NE, Moore MM, et al. Trajectory of mild cognitive impairment onset. *J Int Neuropsychol Soc*. 2008;14(2):192-198.
11. Clem MA, Holliday RP, Pandya S, Hynan LS, Lacritz LH, Woon FL. Predictors that a diagnosis of mild cognitive impairment will remain stable 3 years later. *Cogn Behav Neurol*. 2017;30(1):8-15.
12. Belleville S, Fouquet C, Hudon C, Zomahoun HTV, Croteau J. Consortium for the early identification of Alzheimer's d-Q. neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. *Neuropsychol Rev*. 2017;27(4):328-353.
13. Chan AS, Salmon DP, Butters N, Johnson SA. Semantic network abnormality predicts rate of cognitive decline in patients with probable Alzheimer's disease. *J Int Neuropsychol Soc*. 1995;1(3):297-303.

14. Vaughan RM, Coen RF, Kenny R, Lawlor BA. Preservation of the semantic verbal fluency advantage in a large population-based sample: normative data from the TILDA study. *J Int Neuropsychol Soc.* 2016;22(5):570-576.
15. Canning SJ, Leach L, Stuss D, Ngo L, Black SE. Diagnostic utility of abbreviated fluency measures in Alzheimer's disease and vascular dementia. *Neurology.* 2004;62(4):556-562.
16. Bousfield WA, Sedgewick CHW. An analysis of sequences of restricted associative responses. *J Gen Psychol.* 1944;30(2):149-165.
17. Koenig A, Satt A, Sorin A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's Dement: Diag Assess Dis Monit.* 2015;1:112-124.
18. Chen L, Asgari M, Gale R, Wild K, Dodge H, Kaye J. Improving the assessment of mild cognitive impairment in advanced age with a novel multi-feature automated speech and language analysis of verbal fluency. *Front Psychol.* 2020:11.
19. Eppig J, Wambach D, Nieves C, et al. Dysexecutive functioning in mild cognitive impairment: derailment in temporal gradients. *J Int Neuropsychol Soc.* 2012;18(1):20-28.
20. Meyer D, Messer J, Singh T, et al. Random local temporal structure of category fluency responses. *J Comput Neurosci.* 2011.
21. Lenio S, Lissimore FM, Sajatovic M, et al. Detrending changes the temporal dynamics of a semantic fluency task. *Front Aging Neurosci.* 2016;8(252):1-8.
22. Rohrer JD, Wixted JT, Salmon DP, Butters N. Retrieval from semantic memory and its implications for Alzheimer's disease. *J Exp Psychol Learn Mem Cogn.* 1995;21(5):1127-1139.
23. Lamar M, Price CC, Davis KL, Kaplan E, Libon DJ. Capacity to maintain mental set in dementia. *Neuropsychologia.* 2002;40(4):435-445.
24. Howard VJ. Reasons underlying racial differences in stroke incidence and mortality. *Stroke.* 2013;44:S126-128. 6 Suppl 1.
25. Howard VJ, Cushman M, Pulley L, et al. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology.* 2005;25(3):135-143.
26. Howard VJ, Kleindorfer DO, Judd SE, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. *Ann Neurol.* 2011;69(4):619-627.
27. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. *Med Care.* 2002;40(9):771-781.
28. Howard G, Howard VJ. Twenty years of progress toward understanding the stroke belt. *Stroke.* 2020;51(3):742-750.
29. Shrira I, Christenfeld N, Howard G. Exposure to the US stroke buckle as a risk factor for cerebrovascular mortality. *Neuroepidemiology.* 2008;30(4):229-233.
30. Keyzers C, Gazzola V, Wagenmakers E-J. Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nat Neurosci.* 2020;23(7):788-799.
31. Unverzagt FW, Monahan PO, Moser LR, et al. The Indiana University telephone-based assessment of neuropsychological status: a new method for large scale neuropsychological assessment. *J Int Neuropsychol Soc.* 2007;13(5):799-806.
32. Bowler JV. Vascular cognitive impairment. *J Neurol Neurosurg Psychiatry.* 2005;76:v35-v44. Suppl V.
33. Chan AS, Salmon DP, Butters N, Johnson SA. Semantic network abnormality predicts rate of cognitive decline in patients with Alzheimer's disease. *J Int Neuropsychol Soc: JINS.* 1995;1(3):297-303.
34. Jokinen H, Kalska H, Mantyla R, et al. Cognitive profile of subcortical ischaemic vascular disease. *J Neurol Neurosurg Psychiatry.* 2006;77:28-33.
35. De Looze C, Kelly F, Crosby L, et al. Changes in speech chunking in reading aloud is a marker of mild cognitive impairment and mild-to-moderate Alzheimer's disease. *Curr Alzheimer's Res.* 2018;15(9):828-847.
36. Gayraud F, Barkat-Defradas M, Lee H. Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer's patients and healthy elderly subjects. *J Clin Ling Phon.* 2011;25(3):198-209.
37. Pistono A, Jucla M, Barbeau EJ, et al. Pauses during autobiographical discourse reflect episodic memory process in early Alzheimer's disease. *J Alzheimer's Dis.* 2016;50:687-698.
38. Braak H, Braak E. Neuropathological staging of Alzheimer's-related changes. *Acta Neuropathol.* 1991;82:239-259.
39. Desmond DW. The neuropsychology of vascular cognitive impairment. *J Neurol Sci.* 2004;226:3-7.
40. Damasio H, Tranel D, Grabowski T, Adolphs R, Damasio A. Neural systems behind word and concept retrieval. *Cognition.* 2004;92(1-2):179-229.
41. Garrard P, Lambon Ralph MA, Watson PC, Powis J, Patterson K, Hodges JR. Longitudinal profiles of semantic impairment for living and nonliving concepts in dementia of the Alzheimer's type. *J Cog Neurosci.* 2001;13:892-909.
42. Clark DG, Wadley VG, Kapur P, et al. Lexical factors and cerebral regions influencing verbal fluency performance in MCI. *Neuropsychologia.* 2014;54:98-111.
43. Clark DG, Boan AD, Sims-Robinson C, et al. Differential impact of index stroke on dementia risk in African-Americans compared to whites. *J Stroke Cerebrovasc Dis: Offic J NatStroke Assoc.* 2018;27(10):2725-2730.
44. Kramer JH, Reed BR, Mungas DM, Weiner MW, Chui HC. Executive dysfunction in subcortical ischaemic vascular disease. *J Neurol Neurosurg Psychiatry.* 2002;72:217-220.
45. De Rhonchi D, Palmer K, Pioggiosi P, et al. The combined effect of age, education, and stroke on dementia and cognitive impairment no dementia in the elderly. *Dement Geriatr Cogn Disord.* 2007;24:266-273.
46. Johns EK, Phillips NA, Belleville S, et al. Executive functions in frontotemporal dementia and Lewy body dementia. *Neuropsychology.* 2009;23(6):765-777.
47. Swanberg MM, Tractenberg RE, Mohs R, Thal LJ, Cummings JL. Executive dysfunction in Alzheimer's disease. *Arch Neurol.* 2004;61:556-560.
48. Sudo FK, Amado P, Alves GS, Laks J, Engelhardt E. A continuum of executive function deficits in early subcortical vascular cognitive impairment: a systematic review and meta-analysis. *Dement Neuropsychol.* 2017;11(4):371-380.
49. Crowell TA, Luis CA, Cox DE, Mullan M. Neuropsychological comparison of Alzheimer's disease and dementia with Lewy bodies. *Dement Geriatr Cogn Disord.* 2006;23(2):120-125.
50. De Roeck EE, De Deyn PP, Dierckx E, Engelborghs S. Brief cognitive screening instruments for early detection of Alzheimer's disease: a systematic review. *Alzheimer's Res Ther.* 2019;11(1):21.
51. Meyer DJ, Messer J, Singh T, et al. Random local temporal structure of category fluency responses. *J Comput Neurosci.* 2012;32(2):213-231.
52. Troyer AK, Moscovitch M, Winocur G. Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology.* 1997;11(1):138-146.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ayers MR, Bushnell J, Gao S, et al. Verbal fluency response times predict incident cognitive impairment. *Alzheimer's Dement.* 2022;14:e12277. <https://doi.org/10.1002/dad2.12277>