Journal of Alzheimer's Disease Reports 7 (2023) 1033–1043
DOI 10.3233/ADR-230089
IOS Press

Research Report

Normative Data for Single-Letter Controlled Oral Word Association Test in Older White Australians and Americans, African-Americans, and Hispanic/Latinos

Aoshuang Zhou[a], Carlene Britt[b], Robyn L. Woods[b], Suzanne G. Orchard[b], Anne M. Murray[c,d], Raj C. Shah[e], Ramesh Rajan[f], John J. McNeil[b], Trevor T.-J. Chong[g,h,i], Elsdon Storey[b] and Joanne Ryan[b,*]

[a]*Division of Epidemiology, Jockey Club School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong SAR, China*
[b]*School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia*
[c]*Division of Geriatric and Palliative Medicine, Department of Medicine, Hennepin Healthcare, Minneapolis, MN, USA*
[d]*Berman Center for Outcomes and Clinical Research, Minneapolis, MN, USA*
[e]*Department of Family and Preventive Medicine and the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA*
[f]*Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia*
[g]*Turner Institute for Brain and Mental Health, Monash University, Melbourne, VIC, Australia*
[h]*Department of Neurology, Alfred Health, Melbourne, VIC, Australia*
[i]*Department of Clinical Neurosciences, St Vincent's Hospital, Melbourne, VIC, Australia*

Received 31 July 2023
Accepted 24 August 2023
Pre-press 9 September 2023
Published 19 September 2023

Abstract.
Background: The Controlled Oral Word Association Test (COWAT) is a commonly used measure of verbal fluency. While a normal decline in verbal fluency occurs in late adulthood, significant impairments may indicate brain injury or diseases such as Alzheimer's disease. Normative data is essential to identify when test performance falls below expected levels based on age, gender, and education level.
Objective: This study aimed to establish normative performance data on single-letter COWAT for older community-dwelling adults.
Methods: Over 19,000 healthy men and women, without a diagnosis of dementia or a Modified Mini-Mental State Examination score below 77/100, were recruited for the ASPREE trial. Neuropsychological assessments, including the COWAT with letter F, were administered at study entry.

*Correspondence to: A/Prof. Joanne Ryan, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, 3004, Australia. E-mail: joanne.ryan@monash.edu.

ISSN 2542-4823

Results: Median participant age was 75 years (range 65–98), with 56.5% being women. The majority of participants had 9–11 years of education in Australia and over 12 years in the U.S. The COWAT performance varied across ethno-racial groups and normative data were thus presented separately for 16,335 white Australians, 1,084 white Americans, 896 African-Americans, and 316 Hispanic/Latinos. Women generally outperformed men in the COWAT, except for Hispanic/Latinos. Higher education levels consistently correlated with better COWAT performance across all groups, while the negative association with age was weaker.

Conclusions: This study provides comprehensive normative data for the COWAT stratified by ethno-racial groups in Australia and the U.S., considering age, gender, and education level. These norms can serve as reference standards for screening cognitive impairments in older adults in both clinical and research settings.

Keywords: Aging, Alzheimer's disease, cognitive impairment, normative data, phonemic fluency

INTRODUCTION

Verbal fluency is the ability to express oneself through words and requires interlinked cognitive skills ranging from the retrieval of phonemic and semantic information from memory to selective attention, as well as various elements of executive function [1, 2]. Tests of verbal fluency generally assess an individual's capacity to generate words, either orally or written, within a defined time period while adhering to specific phonemic rules (e.g., starting with a particular letter) or semantic categories (e.g., animals, fruits, vegetables). Therefore, verbal fluency can be broadly classified into two types: phonemic fluency and semantic fluency [3, 4].

Phonemic fluency is widely considered as an indicator of executive function, as generating words within the phonemic category is not a common task in regular speech production, requiring participants to develop non-habitual strategies primarily relying on their lexical knowledge [5, 6]. In general, phonemic fluency increases during childhood and adolescence, followed by a mild decline in older age [7, 8]. However, significant declines are typically indicative of brain injury or diseases [9–11]. Decreased phonemic fluency is observed in individuals with Alzheimer's disease [12, 13], schizophrenia [14, 15], depression [16, 17], Parkinson's disease [18, 19], and after a stroke [20, 21].

The Controlled Oral Word Association Test (COWAT) is a commonly used test of phonemic fluency, with high reliability and validity [22–24]. During the test, participants are required to say all of the words they can think of beginning with a given letter. The testing process engages not only attention regulation and working memory, but also the implementation of strategies for producing word exemplars, monitoring previously produced words, and inhibiting repetitions. It therefore provides a wealth of information regarding an individual's cognitive processes such as language retrieval, cognitive flexibility, and executive functioning [25–27].

Performance on the COWAT is known to be influenced by various factors, including gender, age, education, and ethnicity [28, 29]. It is also sensitive to a range of neuropsychiatric disorders and neurologic conditions [30–32], and has been recognized as an effective tool for detecting brain dysfunction, particularly in relation to abnormalities in the frontal lobes [33, 34]. Indeed, decreased COWAT performance is associated with the early stages of dementia [10, 35]. Furthermore, there is evidence indicating that performance on the COWAT is a significant predictor for the transition from normal cognitive functioning to mild cognitive impairment [35, 36].

Having reliable normative data is crucial to ensure the accuracy of neuropsychological tests which are used as screening tools in clinical practice. It enables clinicians to evaluate the extent to which an individual's performance aligns with the established 'normal' range, thereby identifying potential indications of more serious cognitive impairments that require additional investigation. However, it is important to note that such normative data must consider the diverse characteristics of the population, as well as the various factors that can influence test performance of neurologically intact individuals.

Numerous previous studies on COWAT performance have provided normative data with some sociodemographic variables controlled [37–41]. However, there is currently no large sample of comprehensive norms available across age, gender, and education. Additionally, while many individuals are reaching older age in relatively good health with the increase in life expectancy worldwide [42], there is an escalating need for a specific focus on older adults [25]. Therefore, it is important to

establish high-quality normative data to assist clinicians and researchers in better interpreting their findings for older community-dwelling populations. This necessitates considering not only the influence of education, gender, and age, but also cultural backgrounds.

The purpose of the present study is to establish performance norms for older community-dwelling adults using the COWAT with the single letter F. The normative data is stratified based on age, gender, and education level. Given the importance of cultural differences across racial and ethnic groups, all data are presented separately for white Australians and Americans, African-Americans, and Hispanic/Latinos.

METHODS

Participants

This analysis used cross-sectional data collected at baseline (enrolment) from the ASPirin in Reducing Events in the Elderly (ASPREE) study. ASPREE was a randomized, double-blind, placebo-controlled trial examining the effects of low-dose aspirin in initially healthy older adults [43]. Detailed information, including the study design, has been published previously [44, 45].

Participants were eligible if they were community-dwelling and aged over 70 years (or at least 65 years for U.S. African-American and Hispanic/Latino minority groups) at study enrolment. Cognitive performance was assessed using the Modified Mini-Mental State Examination (3MS) [46], with a score of > 77/100 required for trial inclusion. The exclusion criteria included diagnosed dementia, established cardiovascular disease or past myocardial event, uncontrolled hypertension (SBP > 180 mmHg and /or DBP > 105 mmHg), atrial fibrillation, independence-limiting physical disability, contraindication to aspirin, high risk of bleeding, anemia, and any health condition likely to cause death within 5 years.

ASPREE was conducted in accordance with the Declaration of Helsinki 1964 as revised in 2008, the NHMRC Guidelines on Human Experimentation, the federal patient privacy (HIPAA) law and ICH-GCP guidelines, and the International Conference of Harmonization Guidelines for Good Clinical Practice. It also followed the Code of Federal Regulations as it corresponded to areas of clinical research. The ASPREE Steering Committee held overall responsibility for the conduct of the trial, and all participants in ASPREE provided written informed consent.

Procedure and instruments

ASPREE participants were recruited between 2010 and 2014 from both rural and metropolitan areas in the southeastern states of Australia, as well as multiple centers across the U.S. Eligibility of participants was confirmed after two face-to-face baseline visits, staged one month apart. During these visits, data on neurocognitive assessments, physical examinations, and other baseline questionnaires were collected prior to randomization.

To ensure the accuracy and consistency in the delivery of cognitive measures among all ASPREE staff members, strict training sessions led by study investigators were conducted. Staff members also underwent annual re-accreditation throughout the four-year baseline collection period (2010-2014). A wide range of demographic information was collected for all participants, including age, gender, race, ethnicity (including Caucasian, Aboriginal/Torres Strait Islander, Asian, African American, Hispanic, American Indian), and years of formal education.

Cognitive assessments administered at baseline included the COWAT. While the complete COWAT involves administering from 3 different alphabet prompts, including 'F+A+S' or 'C+F+L', the assessment of verbal fluency in ASPREE involved only a single trial of words beginning with the letter 'F', noting the high internal consistency between different letters [47]. This was done to reduce the length of the overall in-person study assessments, given the age of participants, and the range of other information also collected, and thus ensure that the participants were not overly burdened. Participants were first provided instructions using the letter B as an example: they were asked to orally produce as many words as possible when provided with the letter, and were instructed not to include proper nouns (e.g., using the example letter of B, "such as Boston, Bob or Buick"), or repeat the same word with a different ending (e.g., be, been) [7]. All staff used a compulsory script to minimize variation in participant instructions across the sites. These instructions were based on the guidelines provided in the Multilingual Aphasia Examination manual [48]. Scoring procedures on the COWAT test were standardized, and specific guidance was provided for unusual words that arose during the task. If

the participant recalled similar sounding words, staff queried the participant for their definitions at the time of administration to delineate between a unique similar sounding word (e.g., for, fore and four) versus a repetition. The test was completed within a precisely timed 60-s interval. Therefore, the COWAT score represents the number of valid words produced within 1 min.

Statistical analysis

Mean (SD) scores on the COWAT were calculated for the major ethno-racial groups, namely whites (separately for participants in Australia and the U.S.), African-Americans, and Hispanic/Latinos. Individuals from minor ethno-racial groups ($n = 395$), including Aboriginal and Torres Strait Islanders, Asian, and American Indians, were excluded from current analysis due to insufficient sample sizes. Similarly, those who completed the test in Spanish ($n = 57$) were also excluded due to low sample size.

Multiple linear regression was used to examine the effects of age, gender, education, and ethno-racial status on COWAT performance. Additional linear regression was performed by stratifying the analysis based on ethno-racial groups to further explore these effects. Age was categorized into four groups: 65–69 years (U.S. only), 70–74 years, 75–79 years, and 80+ years. Similar categorization was applied to education attainment. In the Australian cohort, education was divided into five streams:<9 years, 9–11 years, 12 years (indicating completion of high school education), 13–16 years (representing University/Technical and further education), and > 16 years (suggesting post graduate education). Each category had a minimum of 20 participants to ensure sufficient representation. The U.S. subpopulations initially followed similar education categories as the Australian cohort. However, due to inadequate representation at certain education levels, particularly < 12 years, the U.S. subpopulations were combined into two groups based on years of education:<12 years and 12 years or more.

The standard regression model assumptions, such as homogeneity of variance and approximate normal distribution of the residuals, were examined and found to be valid. Two-way interactions between age, gender and education were also examined. Stata version 18.0 (StataCorp, College Station, USA) was used for all statistical analyses.

RESULTS

This study sample comprised 97.5% of participants from the ASPREE trial, including 16,335 Australian whites and 2,296 participants in the U.S., with the latter comprised of 47.2% white, 39.0% African-Americans, and 13.8% Hispanic/Latinos, respectively. Individuals in this sample ranged in age from 65 to 98 years, with a median age of 74 and a mean age of 75. More than half of the participants (56.6%) were women. Education attainment varied across ethno-racial groups. While the majority of participants in the U.S. (70.4%) reported having more than 12 years of education, the highest proportion of participants from Australia reported having 9 to 11 years of education (33.2%). Full details of the baseline characteristics of the ASPREE cohort have been described previously [45].

The COWAT scores across the entire cohort and each ethno-racial category showed approximately normal distributions (Supplementary Figures 1 and 2). Summary statistics for the COWAT scores based on the demographic characteristics of the study participants are shown in Table 1. The overall mean score of COWAT performance was 12.1 (SD 4.6), with a range between 0 and 36. The mean score varied from 10.4 for the lowest education level (<9 years) to 13.6 for individuals with more than 16 years of education.

The results of the multivariable linear regression model for COWAT performance, including all demographic variables in a single model, are presented in Supplementary Table 1. The analysis suggests significant associations between COWAT performance and all variables, indicating that individuals with higher educational level, women, or younger age exhibited better performance. However, it is worth noting that age becomes a significant predictor for COWAT performance only for individuals aged 80 years and above. The COWAT scores varied across the four ethno-racial groups, with best performance observed in Australian whites, followed by U.S whites, Hispanic/Latinos, and African-Americans. Considering the potential influence of cultural differences on performance within these groups, all subsequent analyses and normative data were stratified by ethno-racial group.

Tables 2 and 3 present stratified analyses for Australian and U.S. participants separately. A negative association between age and COWAT performance was observed in the white population from both Australia and the U.S. but not for U.S. minority groups (African-Americans and Hispanic/Latinos). Of note,

Table 1
Summary statistics for the Controlled Oral Word Association Test (COWAT)

Characteristics	N	%	Mean	S.D.	Range
Overall[1]	18,631		12.1	4.6	0 to 36
Ethno-racial groups					
white Australians	16,335	87.7	12.1	4.6	0 to 36
white U.S.	1,084	5.8	12.8	4.4	3 to 30
African American U.S.	896	4.8	11.3	3.9	2 to 25
Hispanic/Latino U.S.	316	1.7	10.6	3.8	1 to 21
Gender					
Female	10,541	56.6	12.6	4.6	0 to 36
Male	8,090	43.4	11.5	4.5	1 to 34
Age, y					
65–69[2]	537	2.9	11.3	4.0	2 to 25
70–74	10,300	55.3	12.3	4.6	0 to 34
75–79	4,904	26.3	12.0	4.5	1 to 33
80–85+	2,890	15.5	11.8	4.7	1 to 36
Education level, y					
<9	2,921	15.7	10.4	4.1	0 to 27
9–11	5,538	29.7	11.5	4.3	1 to 33
12	2,253	12.1	11.9	4.4	1 to 33
13–15	3,169	17.0	12.6	4.6	1 to 33
16+	4,750	25.5	13.6	4.7	1 to 36
<9 to 12 years	10,712	57.5	11.3	4.3	0 to 33
>12 years	7,919	42.5	13.2	4.7	1 to 36

S.D., standard deviation. [1]Not included in these tables are individuals who did not provide their education level ($n = 1$) or did not complete the COWAT ($n = 30$), individuals in another ethno-racial group ($n = 395$), and individuals who completed the COWAT in Spanish ($n = 57$). [2]African American and Hispanic/Latino U.S. participants only, were eligible for inclusion in ASPREE from 65 years of age.

Table 2
Linear regression model of the association between demographic characteristics and Controlled Oral Word Association Test (COWAT) scores of Australian whites ($n = 16,335$)

	β (S.E.)	p
Gender		
Male	Ref.	
Female	1.26 (0.07)	<0.0001
Age, y		
70–74	Ref.	
75–79	–0.17 (0.08)	0.04
80+	–0.32 (0.10)	0.001
Education level, y		
<9	Ref.	
9–11	1.01 (0.10)	<0.0001
12	1.74 (0.13)	<0.0001
13–15	2.23 (0.12)	<0.0001
16+	3.34 (0.14)	<0.0001

however, U.S. minority participants were aged 65+ at enrolment and given the smaller numbers, could only be separated into two groups, while the latter was 70+ years.

Higher levels of education were consistently linked to improved performance on the COWAT in all ethno-racial groups. Likewise, a highly significant association between gender and COWAT scores was observed across all ethno-racial groups, except for Hispanic/Latinos. The results suggest that women generally performed better than men on the COWAT, with a notably stronger effect observed in Australian whites ($\beta = 1.26$) compared to U.S. whites ($\beta = 0.83$) and African-Americans ($\beta = 0.56$).

Normative data for COWAT in Australian whites is shown in Table 4, categorized by five levels of education and three age groups, and presented separately for men and women. Notably, women consistently exhibited higher mean scores compared to men across all sub-categories. Within each age group, COWAT scores increased with more years of education, with only one exception for men aged over 80. Among this group, individuals with 9-10 years of education showed a slightly higher mean score compared to those with 12 years of education. Furthermore, the pattern of declining performance associated with aging was specifically observed in higher educational groups, namely those with 13–15 years and more than 16 years of education.

Normative data for U.S. participants is provided in Table 5, organized by gender categories, and divided into two age and education groups. The data are presented separately for white, African-American, and Hispanic/Latino participants. Given the small number of white U.S. men with less than 12 years of educa-

Table 3
Linear regression model of the association between demographic characteristics and Controlled Oral Word Association Test (COWAT) scores of U.S. participants

	β (S.E.)	*p*
Whites (n = 1,084)		
Gender		
Male	Ref.	
Female	0.83 (0.28)	0.003
Age, y		
70–74	Ref.	
75–79	–0.46 (0.30)	0.12
80+	–0.73 (0.35)	0.04
Education level, y		
≤12	Ref.	
>12	1.85 (0.33)	<0.0001
African-Americans (n = 896)		
Gender		
Male	Ref.	
Female	0.56 (0.27)	0.04
Age, y		
65–69	Ref.	
70+	–0.36 (0.26)	0.17
Education level, y		
≤12	Ref.	
>12	2.09 (0.27)	<0.0001
Hispanic/Latinos (n = 316)		
Gender		
Male	Ref.	
Female	0.69 (0.41)	0.10
Age, y		
65–69	Ref.	
70+	0.06 (0.40)	0.87
Education level, y		
≤12	Ref.	
>12	2.87 (0.41)	<0.0001

tion, we could not present normative data for three age groups, and grouped the 75–79 and 80+ age groups. Similar to the findings in the Australian cohort, women outperformed men across all sub-categories, except for highly educated Hispanic/Latinos aged between 65 and 69 years, where men displayed higher mean scores. Additionally, in the U.S. white subgroup, mean scores were lower in the higher age group, whereas relatively close or equal mean scores were consistently observed in African-Americans and Hispanic/Latinos.

To assess whether participants who developed dementia soon after enrolment affected the normative data, we excluded these who were adjudicated as having dementia within three years of baseline cognitive testing. As can be seen from the Supplementary Table 2, given the small number of incident cases overall, normative data that excludes 217 white Australians with diagnosed dementia in the first three years shows negligible change. The differences in SD are less than 0.1 across all categories, except for white Australian men aged over 80 with 9–11 years of education, where the SD changed by 0.2. Similarly, when excluding U.S. individuals with adjudicated dementia diagnosis within three years of baseline cognitive testing (26 whites, 15 African-Americans, and 3 Hispanic/Latinos), there was no significant change in the normative data, with ≤ 0.1 difference in the SD, except for Hispanic/Latino men aged over 70 with more than 12 years of education, where the SD decreased by 0.2.

DISCUSSION

This study provides new reference values for cognitive performance on the COWAT in a relatively healthy older population from four distinct cultural groups: Australian whites, U.S. whites, African-Americans, and Hispanic/Latinos. The sample included men and women over 70 years of age (65 years for U.S. minorities) residing in communities across the U.S. and southeastern states of Australia. To date, this is the largest sample size used to characterize COWAT (single letter version) normative scores for older populations across differ-

Table 4
Age, education, and gender specific reference values for Controlled Oral Word Association Test (COWAT) scores in Australian whites (*n* = 16,335)

	Education < 9 y		Education 9–11 y		Education 12 y		Education 13–15 y		Education ≥ 16 y	
Age, y	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
70–74	9.7 (3.9) n = 667	11.0 (4.1) n = 774	10.7 (4.2) n = 1,324	12.1 (4.2) n = 1,791	11.8 (4.4) n = 445	12.9 (4.5) n = 559	12.3 (4.6) n = 671	13.6 (4.7) n = 782	13.4 (4.8) n = 1,258	14.6 (4.8) n = 1,158
75–79	9.9 (4.1) n = 356	10.9 (4.1) n = 461	10.9 (4.4) n = 603	12.2 (4.5) n = 893	11.7 (4.4) n = 199	12.5 (4.5) n = 268	11.9 (4.6) n = 252	12.8 (4.4) n = 390	12.9 (4.3) n = 452	14.2 (4.7) n = 461
80+	9.8 (4.3) n = 239	11.2 (4.3) n = 339	11.1 (4.3[1]) n = 289	11.7 (4.5) n = 527	10.9 (4.3) n = 123	12.6 (4.6) n = 176	11.3 (4.6) n = 186	13.1 (4.7) n = 211	12.2 (4.9) n = 257	14.3 (5.1) n = 224

If individuals in the pre-dementia diagnosis phase are excluded (217 individuals with an adjudicated dementia diagnosis within three years of baseline cognitive testing), there is no appreciable change in the normative data (≤0.1 difference in the SD with the exception of [1] where the SD is 4.1 (refer Supplementary Table 2 for complete data).

Table 5
Age, education, and gender specific reference values, mean (SD) for Controlled Oral Word Association Test (COWAT) in U.S. whites ($n = 1{,}084$), African Americans ($n = 896$) and Hispanic/Latinos ($n = 316$)

	Education 0–12 y		Education > 12 y	
	Male	Female	Male	Female
U.S. White				
70–74	10.4 (3.4), $n = 28$	11.9 (4.7), $n = 56$	13.0 (4.2), $n = 131$	13.7 (4.5), $n = 271$
75–85+	9.5 (3.6), $n = 31$	11.8 (4.0), $n = 88$	12.6 (4.6), $n = 155$	13.0 (4.2), $n = 324$
U.S. African American				
65–69	9.6 (3.9), $n = 54$	10.4 (3.7), $n = 64$	11.5 (4.0), $n = 101$	12.5 (3.8), $n = 183$
70–85+	9.3 (3.5), $n = 61$	10.0 (3.5), $n = 106$	11.9 (3.8), $n = 87$	11.9 (4.0), $n = 240$
U.S. Hispanic/Latino				
65–69	8.6 (3.2), $n = 28$	9.9 (3.4), $n = 47$	13.1 (4.3), $n = 23$	12.0 (3.7), $n = 37$
70–85+	8.8 (3.3), $n = 37$	10.0 (3.6), $n = 79$	11.9 (3.3[1]), $n = 33$	12.6 (3.4), $n = 32$

If individuals in the pre-dementia diagnosis phase are excluded (26 white, 15 African-American, and 3 Hispanic/Latino individuals with an adjudicated dementia diagnosis within three years of baseline cognitive testing), there is no appreciable change in the normative data ($\leq$0.1 difference in the SD with the exception of [1]where the SD is 3.1.

ent racial and ethnic groups, and across age, gender, and educational levels. The findings provide practical information to clinicians and researchers as a reference for the older adults, suggesting higher education level, women, and younger age were all associated with better performance. However, some variations in performance were observed among different ethno-racial groups.

The COWAT is a multi-dimensional task that primarily measures phonemic fluency and executive functioning [33, 49]. It is also easily administered and straight forward to score, requiring no lengthy training for standardized administration. An individual's performance on phonemic fluency depends on their stored knowledge, their ability to phonologically search and retrieve relevant information based on that knowledge, and their capacity to maintain selective attention in a sustained manner throughout the task duration [2, 25]. While lower performance on phonemic fluency does not necessarily indicate specific neurological conditions, they can be used as a sensitive marker for detecting cognitive impairment [35, 36]. As such, COWAT is usually used for screening in both clinical and research settings. However, the successful implementation of such screening practices heavily relies on high-quality reference norms.

Over the past few decades, a range of studies have presented normative data for phonemic fluency in general populations, using trials with either single or triple letters [29, 37, 38, 47, 50–60]. These studies have been conducted across various countries, including the U.S. [29, 37, 38], Canada [47, 53], Brazil [52, 55], Spain [56], France [58], Greece [54], Italy [51], Iceland [59], Ireland [60], and so forth [50, 57, 58]. However, most of these studies have recruited relatively small sample sizes, typically involving fewer than 500 individuals [38, 51–57, 61]. This limitation has potentially impeded the development of norms specific to different demographic groups. Furthermore, most of these studies have focused on individuals with a wide age range (e.g., 16–95 years) [37, 47, 51, 53, 54, 56, 57, 59] or specifically targeted younger populations (e.g., children and adolescents) [50, 52, 55]. As a result, the generalization of the conclusions from these studies to older adults may be limited. This further highlights the need for comprehensive normative COWAT data derived from a large-scale sample specifically for older population.

To the best of our knowledge, the previous largest study focusing on normative data for COWAT performance in older English-speaking adults who were cognitively healthy included only 743 participants aged between 56 and 97 years [29]. Consistent with the findings of the current study, they reported that COWAT performance was most strongly influenced by education, with relatively mild negative effects observed for age. There were gender differences, although a bit smaller than those observed for age, and they did not present gender-adjusted norms. Additionally, as their sample predominantly consisted of Caucasians, they were unable to investigate the effects of race and ethnicity. This limitation was common in most studies that exclusively included a single ethno-racial category [38, 47, 51, 62], which may have limited the direct comparison between studies due to potential variations in demographics across different populations, as well as analytical methods.

Supporting the findings of our study, education has been consistently associated with improved performance on the COWAT [56, 58, 63–66], across

different countries and ethnicities. This includes studies of 309 African Americans (with years of education ranging from 0–8 years to 16–20 years) [66], 257 Spanish (with years of education categorized into two levels:<13,≥13 years) [56], and 1133 French (with education attainment categorized into low to high groups based on years of education and diploma achievement) [58]. We also found that education had the strongest association with test performance, compared to other key demographic factors such as age and gender in line with prior work [56, 62, 67, 68]. In addition, our study revealed an inverse trend between the proportion of higher education level and the magnitude of education influence within the U.S. ethno-racial groups. For example, while Hispanic/Latinos reported the lowest proportion of having more than 12 years of education (39.6%), the influence of education was the greatest ($\beta = 2.87$), compared to other U.S. populations.

A comprehensive meta-analysis of 26 studies involving more than 21,000 individuals with an age range of 15 to 95 years by Rodriguez-Aranda et al. [69] revealed an age-dependent decline in COWAT performance after 40 years of age, which accelerated after the age of 60 years. In contrast, we found that only after the age of 80 years was a significant decline in COWAT performance observed. This could be explained by the differences in the baseline cognitive status of study participants, as our findings specifically relate to individuals without major cognitive impairment, and who were in relatively good health to participate in the study. In the meta-analysis, 11 out of the 26 studies [69] did not screen for dementia. It is, therefore, reasonable to observe a decline at a later age in our cognitively intact sample. Additionally, considering our ASPREE sample tended to be well educated (median years of education was 12), the observed late age-related differences may also result from the protective effects of education, which may have the potential to attenuate the negative effects of aging in individuals with higher education level [56].

In our study, women performed significantly better on the COWAT than men across all ethno-racial groups, except for Hispanic/Latinos, where the difference was not statistically significant. These findings largely align with previous investigational studies [3, 70–72] and the aforementioned meta-analytical study [69] that reported gender difference favoring women. Of note however, the number of Hispanic/Latino individuals in the ASPREE cohort was relatively small ($n = 316$), compared to white Australians ($n = 16{,}335$), white U.S. residents ($n = 1{,}084$), and African-Americans ($n = 896$). Additionally, Hispanic/Latino individuals in ASPREE reported lower levels of education in comparison to the other ethno-racial groups [45]. It could be inferred that the insignificant difference observed in Hispanic/Latino group may be because the effects of education are larger than gender, thus a significant enough gender difference could only be observed in sufficiently large sample sizes. This further highlights the necessity for larger samples in future studies to provide more comprehensive normative data for Hispanic/Latinos.

Strengths and limitations

The primary limitation of our study is that older adults recruited to the ASPREE trial were generally healthy and thus may not fully represent all older community-dwelling individuals. Another limitation is the underrepresentation of other ethno-racial minorities, including Aboriginal and Torres Strait Islanders, Asians, and American Indians, which will limit the generalizability of the normative data. Additionally, a limitation is the difference in staff who administered the neurological assessments across different sites. To mitigate this, extensive initial training was provided, and annual re-accreditation by neuropsychologists was implemented. Staff members who did not meet the qualification criteria in quality control underwent retraining. Finally, our normative data only focused on phonemic fluency of the single letter F, and did not include semantic fluency. This was based on overall feasibility and time constraints of each visit, considering the large scale of the study. Future studies are thus warranted with inclusion of both phonemic (FAS) and semantic fluency to provide more comprehensive reference norms.

The strengths of our study include the large size of the ASPREE sample, the inclusion of different ethno-racial groups, and the standardized administration of the tests. While the sample sizes of African-Americans and Hispanic/Latinos are notably smaller compared to the proportion of whites, they are still relatively larger than those in previous studies. In fact, this study represents the largest investigation providing normative data for COWAT performance in the older population, and presented data separately for white Australians, white Americans, African-Americans, and Hispanic/Latinos, with stratification by age, gender, and education. This study will also provide clinicians and researchers with useful resources to select appropriate reference standards

that best suit the demographic characteristics of their targeted individuals.

Conclusion

This study is the first, and the largest, to characterize normative data for COWAT performance in older adults according to key demographics across four ethno-racial groups. The findings show that education, gender, and age are all independent factors that influence performance on the COWAT, with varying associations across ethno-racial groups. The new reference value presented in this study makes a significant contribution to the field and will facilitate the application of the COWAT in both clinical and research settings for screening older adults with low or atypical cognitive function. These findings emphasize the importance of using appropriate normative data that aligns with the population under examination. Moreover, careful consideration should be given to adjusting for socio-demographic factors in future epidemiological studies focuses on cognitive aging.

ACKNOWLEDGMENTS

The authors acknowledge the dedicated and skilled staff in Australia and the Unites States for the conduct of the trial. The authors also are most grateful to the ASPREE participants, who so willingly volunteered for this study, and the general practitioners and medical clinics who support the participants in the ASPREE study. Trial Registration: International Standard Randomized Controlled Trial Number Register (ISRCTN83772183) and clinicaltrials.gov (NCT01038583).

FUNDING

The work was supported by the National Institute on Aging and the National Cancer Institute at the National Institutes of Health (grant number U01AG029824); the National Health and Medical Research Council (NHMRC) of Australia (grant numbers 334047 and 1127060); Monash University (Australia); the Victorian Cancer Agency (Australia). JR is funded by a NHMRC Research Leader Fellowship (2016438).

CONFLICT OF INTEREST

The authors report no conflicts of interest.

DATA AVAILABILITY

The data support the findings of this study are available within the article and its supplementary material.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/ADR-230089.

REFERENCES

[1] Shao Z, Janse E, Visser K, Meyer AS (2014) What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Front Psychol* **5**, 772.

[2] Troyer AK, Moscovitch M, Winocur G (1997) Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology* **11**, 138-146.

[3] Crossley M, D'Arcy C, Rawson NS (1997) Letter and category fluency in community-dwelling Canadian seniors: A comparison of normal participants to those with dementia of the Alzheimer or vascular type. *J Clin Exp Neuropsychol* **19**, 52-62.

[4] Harrison JE, Buxton P, Husain M, Wise R (2000) Short test of semantic and phonological fluency: Normal performance, validity and test-retest reliability. *Br J Clin Psychol* **39**, 181-191.

[5] Luo L, Luk G, Bialystok E (2010) Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition* **114**, 29-41.

[6] Henry JD, Crawford JR (2004) A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology* **18**, 284-295.

[7] Strauss E, Sherman EMS, Spreen O (2006) *A compendium of neuropsychological tests: Administration, norms and commentary (3rd ed.)*, Oxford University Press, New York.

[8] Elgamal SA, Roy EA, Sharratt MT (2011) Age and verbal fluency: The mediating effect of speed of processing. *Can Geriatr J* **14**, 66-72.

[9] Batty R, Francis A, Thomas N, Hopwood M, Ponsford J, Johnston L, Rossell S (2015) Verbal fluency, clustering, and switching in patients with psychosis following traumatic brain injury (PFTBI). *Psychiatry Res* **227**, 152-159.

[10] Henry JD, Crawford JR, Phillips LH (2004) Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia* **42**, 1212-1222.

[11] Zaninotto AL, de Paula Guirado VM, Baldivia B, Núñes MD, Amorim RL, Teixeira MJ, de Lucia MC, de Andrade AF, Paiva WS (2014) Improvement of verbal fluency in patients with diffuse brain injury over time. *Neuropsychiatr Dis Treat* **10**, 1155-1160.

[12] Braaten AJ, Parsons TD, McCue R, Sellers A, Burns WJ (2006) Neurocognitive differential diagnosis of dementing diseases: Alzheimer's dementia, vascular dementia, frontotemporal dementia, and major depressive disorder. *Int J Neurosci* **116**, 1271-1293.

[13] Laws KR, Duncan A, Gale TM (2010) 'Normal' semantic-phonemic fluency discrepancy in Alzheimer's disease? A meta-analytic study. *Cortex* **46**, 595-601.

[14] Brébion G, Stephan-Otto C, Ochoa S, Nieto L, Contel M, Usall J (2018) Verbal fluency in male and female schizophrenia patients: Different patterns of association with processing speed, working memory span, and clinical symptoms. *Neuropsychology* **32**, 65-76.

[15] Kremen WS, Seidman LJ, Faraone SV, Tsuang MT (2003) Is there disproportionate impairment in semantic or phonemic fluency in schizophrenia? *J Int Neuropsychol Soc* **9**, 79-88.

[16] Akiyama T, Koeda M, Okubo Y, Kimura M (2018) Hypofunction of left dorsolateral prefrontal cortex in depression during verbal fluency task: A multi-channel near-infrared spectroscopy study. *J Affect Disord* **231**, 83-90.

[17] Henry J, Crawford JR (2005) A meta-analytic review of verbal fluency deficits in depression. *J Clin Exp Neuropsychol* **27**, 78-101.

[18] Lim KB, Kim J, Lee HJ, Yoo J, Kim HS, Kim C, Lee H (2019) COWAT performance of persons with Alzheimer dementia, vascular dementia, and Parkinson disease dementia according to stage of cognitive impairment. *PM R* **11**, 737-744.

[19] Torralva T, Laffaye T, Báez S, Gleichgerrcht E, Bruno D, Chade A, Ibañez A, Manes F, Gershanik O, Roca M (2015) Verbal fluency as a rapid screening test for cognitive impairment in early Parkinson's disease. *J Neuropsychiatry Clin Neurosci* **27**, 244-247.

[20] Schmidt CSM, Nitschke K, Bormann T, Römer P, Kümmerer D, Martin M, Umarova RM, Leonhart R, Egger K, Dressing A, Musso M, Willmes K, Weiller C, Kaller CP (2019) Dissociating frontal and temporal correlates of phonological and semantic fluency in a large sample of left hemisphere stroke patients. *Neuroimage Clin* **23**, 101840.

[21] Sarno MT, Postman WA, Cho YS, Norman RG (2005) Evolution of phonemic word fluency performance in post-stroke aphasia. *J Commun Disord* **38**, 83-107.

[22] Abe M, Suzuki K, Okada K, Miura R, Fujii T, Etsurou M, Yamadori A (2004) [Normative data on tests for frontal lobe functions: Trail Making Test, Verbal fluency, Wisconsin Card Sorting Test (Keio version)]. *No To Shinkei* **56**, 567-574.

[23] L. Benton KdH, Abigail B. Sivan (1983) *Multilingual aphasia examination (3rd ed.)*, AJA Associates, Iowa City, IA.

[24] Ross TP (2003) The reliability of cluster and switch scores for the Controlled Oral Word Association Test. *Arch Clin Neuropsychol* **18**, 153-164.

[25] Villalobos D, Torres-Simón L, Pacios J, Paúl N, Del Río D (2022) A systematic review of normative data for verbal fluency test in different languages. *Neuropsychol Rev*, doi: 10.1007/s11065-022-09549-0.

[26] Reitan RM, Wolfson D (1993) *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation*, Neuropsychology Press., Dallas, TX.

[27] Abwender DA, Swan JG, Bowerman JT, Connolly SW (2001) Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment* **8**, 323-338.

[28] Bolla KI, Lindgren KN, Bonaccorsy C, Bleecker ML (1990) Predictors of verbal fluency (FAS) in the healthy elderly. *J Clin Psychol* **46**, 623-628.

[29] Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC (1996) Neuropsychological tests' norms above age 55: COWAT, BNT, MAE Token, WRAT-R Reading, AMNART, STROOP, TMT, and JLO. *Clin Neuropsychol* **10**, 262-278.

[30] Iverson GL, Franzen MD, Lovell MR (1999) Normative comparisons for the controlled oral word association test following acute traumatic brain injury. *Clin Neuropsychol* **13**, 437-441.

[31] Hedman E, Hartelius L, Saldert C (2022) Word-finding difficulties in Parkinson's disease: Complex verbal fluency, executive functions and other influencing factors. *Int J Lang Commun Disord* **57**, 565-577.

[32] Engstad T, Almkvist O, Viitanen M, Arnesen E (2003) Impaired motor speed, visuospatial episodic memory and verbal fluency characterize cognition in long-term stroke survivors: The Tromsø Study. *Neuroepidemiology* **22**, 326-331.

[33] Lezak MD, Howieson DB, Loring DW (2004) *Neuropsychological assessment (4th ed.)*, Oxford University Press, New York.

[34] Troyer AK, Moscovitch M, Winocur G, Alexander MP, Stuss D (1998) Clustering and switching on verbal fluency: The effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia* **36**, 499-504.

[35] Frankenberg C, Weiner J, Knebel M, Abulimiti A, Toro P, Herold CJ, Schultz T, Schröder J (2021) Verbal fluency in normal aging and cognitive decline: Results of a longitudinal study. *Comput Speech Lang* **68**, 101195.

[36] McDonnell M, Dill L, Panos S, Amano S, Brown W, Giurgius S, Small G, Miller K (2020) Verbal fluency as a screening tool for mild cognitive impairment. *Int Psychogeriatr* **32**, 1055-1062.

[37] Gladsjo JA, Schuman CC, Evans JD, Peavy GM, Miller SW, Heaton RK (1999) Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment* **6**, 147-178.

[38] Bolla KI, Gray S, Resnick SM, Galante R, Kawas C (1998) Category and letter fluency in highly educated older adults. *Clin Neuropsychol* **12**, 330-338.

[39] Ravdin LD, Katzen HL, Agrawal P, Relkin NR (2003) Letter and semantic fluency in older adults: Effects of mild depressive symptoms and age-stratified normative data. *Clin Neuropsychol* **17**, 195-202.

[40] Fine EM, Kramer JH, Lui LY, Yaffe K (2012) Normative data in women aged 85 and older: Verbal fluency, digit span, and the CVLT-II short form. *Clin Neuropsychol* **26**, 18-30.

[41] Carvalho GA, Caramelli P (2020) Normative data for middle-aged Brazilians in Verbal Fluency (animals and FAS), Trail Making Test (TMT) and Clock Drawing Test (CDT). *Dement Neuropsychol* **14**, 14-23.

[42] Christensen K, Doblhammer G, Rau R, Vaupel JW (2009) Ageing populations: The challenges ahead. *Lancet* **374**, 1196-1208.

[43] McNeil JJ, Woods RL, Nelson MR, Reid CM, Kirpach B, Wolfe R, Storey E, Shah RC, Lockery JE, Tonkin AM, Newman AB, Williamson JD, Margolis KL, Ernst ME, Abhayaratna WP, Stocks N, Fitzgerald SM, Orchard SG, Trevaks RE, Beilin LJ, Donnan GA, Gibbs P, Johnston CI, Ryan J, Radziszewska B, Grimm R, Murray AM (2018) Effect of aspirin on disability-free survival in the healthy elderly. *N Engl J Med* **379**, 1499-1508.

[44] ASPREE IG (2013) Study design of ASPirin in Reducing Events in the Elderly (ASPREE): A randomized, controlled trial. *Contemp Clin Trials* **36**, 555-564.

[45] McNeil JJ, Woods RL, Nelson MR, Murray AM, Reid CM, Kirpach B, Storey E, Shah RC, Wolfe RS, Tonkin AM, Newman AB, Williamson JD, Lockery JE, Margolis KL, Ernst ME, Abhayaratna WP, Stocks N, Fitzgerald SM, Trevaks RE, Orchard SG, Beilin LJ, Donnan GA, Gibbs P, Johnston CI, Grimm RH (2017) Baseline characteristics of participants in the ASPREE (ASPirin in Reducing Events

in the Elderly) Study. *J Gerontol A Biol Sci Med Sci* **72**, 1586-1593.

[46] Teng EL, Chui HC (1987) The Modified Mini-Mental State (3MS) examination. *J Clin Psychiatry* **48**, 314-318.

[47] Tombaugh TN, Kozak J, Rees L (1999) Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Arch Clin Neuropsychol* **14**, 167-177.

[48] Benton AL, Sivan AB, Hamsher K deS, Varney NR, Spreen O (1994) *Contributions to neuropsychological assessment: A clinical manual (2nd Ed)*, Oxford University Press, New York.

[49] Ross TP, Furr AE, Carter SE, Weinberg M (2006) The psychometric equivalence of two alternate forms of the Controlled Oral Word Association Test. *Clin Neuropsychol* **20**, 414-431.

[50] Olabarrieta-Landa L, Rivera D, Lara L, Rute-Pérez S, Rodríguez-Lorenzana A, Galarza-Del-Angel J, Peñalver Guia AI, Ferrer-Cascales R, Velázquez-Cardoso J, Campos Varillas AI, Ramos-Usuga D, Chino-Vilca B, Aguilar Uriarte MA, Martín-Lobo P, García de la Cadena C, Postigo-Alonso B, Romero-García I, Rabago Barajas BV, Irías Escher MJ, Arango-Lasprilla JC (2017) Verbal fluency tests: Normative data for Spanish-speaking pediatric population. *NeuroRehabilitation* **41**, 673-686.

[51] Costa A, Bagoj E, Monaco M, Zabberoni S, De Rosa S, Papantonio AM, Mundi C, Caltagirone C, Carlesimo GA (2014) Standardization and normative data obtained in the Italian population for a new verbal fluency instrument, the phonemic/semantic alternate fluency test. *Neurol Sci* **35**, 365-372.

[52] Becker N, Piccolo LR, Salles JF (2019) Verbal fluency development across childhood: Normative data from Brazilian-Portuguese speakers and underlying cognitive processes. *Arch Clin Neuropsychol* **34**, 1217-1231.

[53] Troyer AK (2000) Normative data for clustering and switching on verbal fluency tasks. *J Clin Exp Neuropsychol* **22**, 370-378.

[54] Kosmidis MH, Vlahou CH, Panagiotaki P, Kiosseoglou G (2004) The verbal fluency task in the Greek population: Normative data, and clustering and switching strategies. *J Int Neuropsychol Soc* **10**, 164-172.

[55] Hazin I, Leite G, Oliveira RM, Alencar JC, Fichman HC, Marques PD, de Mello CB (2016) Brazilian normative data on letter and category fluency tasks: Effects of gender, age, and geopolitical region. *Front Psychol* **7**, 684.

[56] Lubrini G, Periáñez JA, Laseca-Zaballa G, Bernabéu-Brotons E, Ríos-Lago M (2022) Verbal fluency tasks: Influence of age, gender, and education and normative data for the Spanish native adult population. *Arch Clin Neuropsychol* **37**, 365-375.

[57] Marquine MJ, Morlett Paredes A, Madriaga C, Blumstein Y, Umlauf A, Kamalyan L, Rivera Mindt M, Suarez P, Artiola IFL, Heaton RK, Cherner M (2021) Demographically-adjusted norms for selected tests of verbal fluency: Results from the Neuropsychological Norms for the US-Mexico Border Region in Spanish (NP-NUMBRS) project. *Clin Neuropsychol* **35**, 269-292.

[58] Auriacombe S, Fabrigoule C, Lafont S, Jacqmin-Gadda H, Dartigues J-F (2001) Letter and category fluency in normal elderly participants: A population-based study. *Aging Neuropsychol Cogn* **8**, 98-108.

[59] Magnusdottir BB, Haraldsson HM, Sigurdsson E (2021) Trail making test, Stroop, and verbal fluency: Regression-based norms for the Icelandic population. *Arch Clin Neuropsychol* **36**, 253-266.

[60] Vaughan RM, Coen RF, Kenny R, Lawlor BA (2016) Preservation of the semantic verbal fluency advantage in a large population-based sample: Normative data from the TILDA Study. *J Int Neuropsychol Soc* **22**, 570-576.

[61] Peña-Casanova J, Quiñones-Ubeda S, Quintana-Aparicio M, Aguilar M, Badenes D, Molinuevo JL, Torner L, Robles A, Barquero MS, Villanueva C, Antúnez C, Martínez-Parra C, Frank-García A, Sanz A, Fernández M, Alfonso V, Sol JM, Blesa R (2009) Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for verbal span, visuospatial span, letter and number sequencing, trail making test, and symbol digit modalities test. *Arch Clin Neuropsychol* **24**, 321-341.

[62] Kim BJ, Lee CS, Oh BH, Hong CH, Lee KS, Son SJ, Han C, Park MH, Jeong HG, Kim TH, Park JH, Kim KW (2013) A normative study of lexical verbal fluency in an educationally-diverse elderly population. *Psychiatry Investig* **10**, 346-351.

[63] O'Bryant SE, Edwards M, Johnson L, Hall J, Gamboa A, O'jile J (2018) Texas Mexican American adult normative studies: Normative data for commonly used clinical neuropsychological measures for English- and Spanish-speakers. *Dev Neuropsychol* **43**, 1-26.

[64] Abdel Aziz K, Khater MS, Emara T, Tawfik HM, Rasheedy D, Mohammedin AS, Tolba MF, El-Gabry DA, Qassem T (2017) Effects of age, education, and gender on verbal fluency in healthy adult Arabic-speakers in Egypt. *Appl Neuropsychol Adult* **24**, 331-341.

[65] Pereira AH, Gonçalves AB, Holz M, Gonçalves HA, Kochhann R, Joanette Y, Zimmermann N, Fonseca RP (2018) Influence of age and education on the processing of clustering and switching in verbal fluency tasks. *Dement Neuropsychol* **12**, 360-367.

[66] Lucas JA, Ivnik RJ, Smith GE, Ferman TJ, Willis FB, Petersen RC, Graff-Radford NR (2005) Mayo's Older African Americans Normative Studies: Norms for Boston Naming Test, Controlled Oral Word Association, Category Fluency, Animal Naming, Token Test, WRAT-3 Reading, Trail Making Test, Stroop Test, and Judgment of Line Orientation. *Clin Neuropsychol* **19**, 243-269.

[67] Esteves CS, Oliveira CR, Moret-Tatay C, Navarro-Pardo E, Carli GAD, Silva IG, Irigaray TQ, Argimon IIdL (2015) Phonemic and semantic verbal fluency tasks: Normative data for elderly Brazilians. *Psicol Reflex Crít* **28**, doi: 10.1590/1678-7153.201528215

[68] Mathuranath PS, George A, Cherian PJ, Alexander A, Sarma SG, Sarma PS (2003) Effects of age, education and gender on verbal fluency. *J Clin Exp Neuropsychol* **25**, 1057-1064.

[69] Rodriguez-Aranda C, Martinussen M (2006) Age-related differences in performance of phonemic verbal fluency measured by Controlled Oral Word Association Task (COWAT): A meta-analytic study. *Dev Neuropsychol* **30**, 697-717.

[70] Capitani E, Laiacona M, Basso A (1998) Phonetically cued word-fluency, gender differences and aging: A reappraisal. *Cortex* **34**, 779-783.

[71] Veroff AE (1980) The neuropsychology of aging. Qualitative analysis of visual reproductions. *Psychol Res* **41**, 2590-2568.

[72] Loonstra AS, Tarlow AR, Sellers AH (2001) COWAT metanorms across age, education, and gender. *Appl Neuropsychol* **8**, 161-166.