

## RESEARCH ARTICLE

# Establishing cognitive baseline in three generations: Framingham Heart Study

Jin Yang<sup>1</sup> | Ting Fang Alvin Ang<sup>2,3</sup> | Sophia Lu<sup>4</sup> | Xue Liu<sup>4</sup> | Sherral Devine<sup>2,3</sup> | Rhoda Au<sup>2,3</sup> | Chunyu Liu<sup>3,4</sup>

<sup>1</sup>Department for Endemic Disease Control and Prevention, Henan Provincial Center for Disease Control and Prevention, Zhengzhou, China

<sup>2</sup>Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, Massachusetts, USA

<sup>3</sup>Framingham Heart Study, Boston University School of Medicine, Boston, Massachusetts, USA

<sup>4</sup>Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA

**Correspondence**

Chunyu Liu, Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA.  
E-mail: [liuc@bu.edu](mailto:liuc@bu.edu)

**Funding information**

FHS, Grant/Award Numbers: N01-HC-25195, HHSN268201500001; National Institute on Aging, Grant/Award Numbers: R01AG059727, R01AG016495, R01AG008122, R01AG062109, U19 AG068753

**Abstract**

**Introduction:** Generational changes warrant recalibrating normative cognitive measures to detect changes indicative of dementia risk within each generation.

**Methods:** We performed linear regressions to compare eight neuropsychological (NP) tests among three-generation cohorts at baseline in Framingham Heart Study (FHS,  $n = 4787$ ) and conducted Cox regressions to investigate the relationships of NP tests with generation-specific dementia risk.

**Results:** The FHS second and third generations performed better than the first generation for seven NP tests (0.14–0.81 standard deviation improvement,  $P \leq .001$ ) while the second and third generations performed similarly for six of eight NP tests ( $P > .05$ ). One standard deviation better performance was associated with a higher reduction in incident dementia risk in the second than the first generation (35% vs. 24%,  $P_{\text{interaction}} = .02$ ) for the similarities test.

**Discussion:** Our findings suggest cohort-based norms are needed for cognitive assessment for the diagnosis of cognitive impairment and dementia.

**KEYWORDS**

Alzheimer's disease, cognitive function, dementia, neuropsychological test, norms

## 1 | INTRODUCTION

A growing body of evidence has shown that age-specific dementia incidence rates are declining for people born later in the twentieth century.<sup>1–3</sup> A recent study also reported that the 5-year age- and sex-adjusted incidence of dementia has declined among the participants of the Framingham Heart Study (FHS) over the course of three decades.<sup>4</sup> The factors contributing to such decline are not completely understood. Impaired cognition is the most important clinical manifestation of dementia, and therefore, monitoring the cognitive changes over time across generations of the same population may provide insight into the different incidence rates of dementia over the past several decades.

However, research to investigate cognitive differences across multiple generations is lacking.

Neuropsychological (NP) tests are used to measure cognitive changes over time.<sup>5</sup> NP tests are often performed under standard conditions and they are scored by comparing to normative standards.<sup>6</sup> However, tremendous improvements in socioeconomic, medical, and lifestyle factors over the past century have led to a generally more cognitively stimulating environment, contributing to the Flynn effect that was first reported in 1984.<sup>7</sup> The Flynn effect refers to the observation that the intelligence quotient (IQ) test scores increased in many parts of the world over the twentieth century.<sup>7</sup> This, in turn, resulted in the rapid improvement in cognitive test scores observed in

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* published by Wiley Periodicals, LLC on behalf of Alzheimer's Association.

several European countries<sup>7-9</sup> and also in the United States in the last several decades.<sup>10,11</sup> Therefore, there is a continuing need for evaluating normative data due to birth cohort effects that are associated with the evolving changes. The FHS is unique for its large, community-based, and ongoing longitudinal surveillance of clinical and cognitive assessment in multiple generations over decades. A similar baseline NP test was administered to the FHS Original cohort,<sup>12-14</sup> the Offspring cohort,<sup>15-17</sup> and the Third Generation cohort<sup>18</sup> using standard NP test administration and scoring procedures over 30 years from the late twentieth century to the early twenty-first century when tremendous changes occurred in human society. Therefore, the multi-generational FHS enables us to evaluate the effects of evolving changes on NP normative assessments across the birth cohorts from the same community-based population.

To that end, the first aim of this study was to compare the common NP tests that were conducted over 30 years across three FHS generations. We also investigated the relationships of the NP tests with generation-specific dementia risk. Our findings may provide complementary information to the observation of the reduction of dementia incidence rate in recent years.

## 2 | METHODS

### 2.1 | Participants

This study identified participants from the three FHS generation cohorts, including the Original cohort (Gen 1,  $n = 5209$ ) enrolled in 1948,<sup>19</sup> the Offspring cohort (Gen 2,  $n = 5214$ ) enrolled in 1971,<sup>20</sup> and the Third Generation cohort ( $n = 4095$ , Gen 3) recruited in 2002.<sup>21</sup> The present study included participants having the baseline NP test from the three generation cohorts (see next section). Age-related cognitive decline is likely to begin before age 60 in healthy educated adults.<sup>9</sup> Therefore, we selected participants aged 55 or older from the three generations at their baseline NP tests. We excluded participants who developed dementia and who had a stroke before the baseline NP tests.<sup>22,23</sup> We also excluded participants with missing education information (Figure 1). Our final sample included 4787 participants ( $n = 2012$  in Gen 1;  $n = 2187$  in Gen 2; and  $n = 588$  in Gen 3; Figure 1). All participants provided written informed consent. The institutional review board at Boston University School of Medicine approved the study procedures.

### 2.2 | Neuropsychological test

At baseline, a brief NP test protocol with eight NP measures was administered to Gen 1 ( $n = 2123$ ) in 1976. A consistent NP test administration and scoring procedures were applied to Gen 2 ( $n = 3022$ )<sup>15-17</sup> starting in 1999, and Gen 3 ( $n = 2420$ )<sup>18</sup> starting in 2009. For this present study, we limited the comparison of the cognitive performance to the eight original NP tests<sup>12</sup> that were commonly administered to

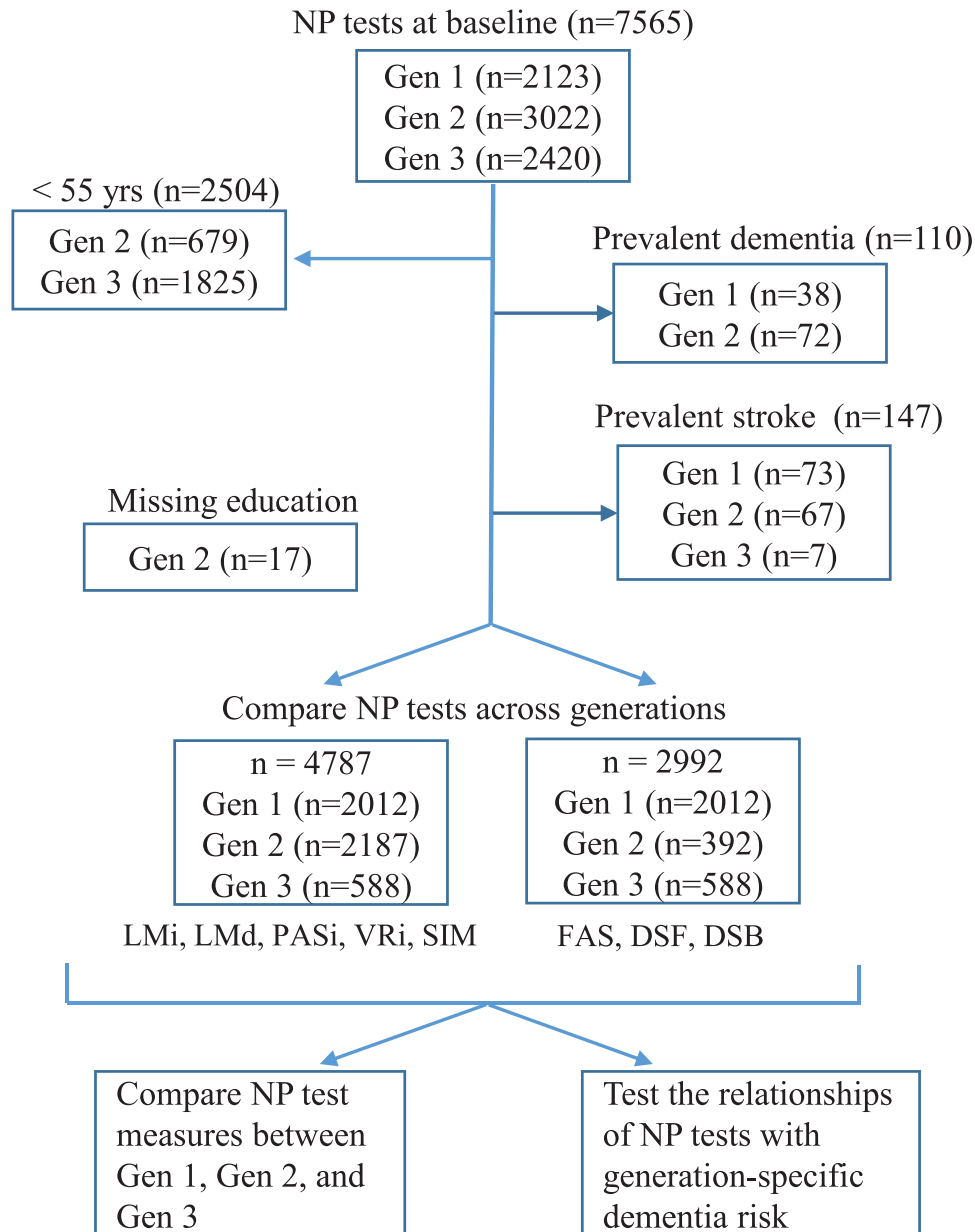
#### RESEARCH IN CONTEXT

- 1. Systematic review:** The authors reviewed the literature using traditional (e.g., PubMed) sources and meeting abstracts and presentations. Over the past century, tremendous improvements in socioeconomic, medical, and lifestyle factors have led to a generally more cognitively stimulating environment. This, in turn, resulted in the rapid increase in cognitive test scores observed in several European countries during the 20th century. Therefore, there is a continuing need for evaluating normative data due to birth cohort effects that are associated with the evolving changes.
- 2. Interpretation:** The findings also suggest that generation-based or cohort-based norms should be applied to cognitive assessment for the diagnosis of cognitive impairment and dementia.
- 3. Future directions:** Our findings have illustrated the need to recognize that, in cognitive assessment, the normative scores to determine significant cognitive changes should be tailored to each generation.

all three cohorts at their baseline assessment (Table S1 in supporting information).

These eight NP tests included the Wechsler Memory Scale (WMS),<sup>24</sup> the Wechsler Adult Intelligence Scale (WAIS),<sup>25</sup> and the Aphasia Examination.<sup>26</sup> Four tests were a part of the WMS, including (1) Logical Memory at Immediate Recall (LMi), (2) Logical Memory at Delayed Recall (LMd), (3) Paired Associate Learning at Immediate Recall (PASi), and (4) Visual Reproduction at Immediate Recall (VRi). The four tests (LMi, LMd, PASi, and VRi) belong to the memory function. Three NP tests were part of the WAIS, including (1) Similarities test (SIM), (2) Digit Span Forward (DSF), and (3) Digit Span Backward (DSB). These three NP tests (SIM, DSF, and DSB) are within the attention and executive function domain. The F, A, and S Letter Fluency (FAS) test came from the Aphasia Examination, and it is the only NP test measured in the verbal fluency domain. Details regarding the assessment of each NP test have been previously described<sup>27</sup> (Table S1).

The sample sizes of the eight NP tests were different due to different sample sizes of these measures in Gen 2. Five tests, LMi, LMd, PASi, VRi, and SIM, were measured in 2187 Gen 2 participants while DSF, DSB, and FAS were measured in 392 Gen 2 participants because these three measures were re-introduced to the standard study protocol in 2005. These 392 Gen 2 participants had similar characteristics compared to the 2187 Gen 2 participants (Table 1; Table S2 in supporting information). To maximize the sample size, we compared LMi, LMd, PASi, VRi, and SIM among generations in 4787 participants, and compared DSF, DSB, and FAS among generations in 2992 participants (Figure 1).



**FIGURE 1** Study design. This study was performed in the Framingham Heart Study Original cohort (Gen 1), Offspring cohort (Gen 2), and Generation 3 cohort (Gen 3). DSB, Digit Span Backward; DSF, Digit Span Forward; FAS, F, A, and S Letter Fluency; LMd, Logical Memory at Delayed Recall; LMi, Logical Memory at Immediate Recall; NP test, neuropsychological test; PASi, Paired Associate Learning (Immediate Recall); SIM, Similarities; VRi, Visual Reproduction at Immediate Recall.

## 2.3 | Surveillance for dementia

The cognitive status of all FHS participants has been closely monitored using consistent dementia surveillance criteria. When possible, participants with low cognitive scores underwent a neurologic assessment.<sup>12</sup> The dementia diagnostic panel included at least one neuropsychologist and one neurologist. This panel reviewed and determined whether a person had dementia, the date of dementia onset, and the dementia subtype<sup>4,12</sup> based on standard criteria from the National Institute of Neurological and Communicative

Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association.<sup>23,28</sup>

## 2.4 | Statistical analyses

### 2.4.1 | Standardization of NP scores

We obtained the mean and standard deviation (SD) of each of the eight NP tests in the Gen 1 participants. We then obtained the Z scores

**TABLE 1** Characteristics of the Framingham Heart Study participants at NP test baseline.

Characteristic	Gen 1 (n = 2012)	Gen 2 (n = 2187)	Gen 3 (n = 588)
Birth year (Mean ± SD)	1909.8 ± 5.7	1932.8 ± 7.6	1949.4 ± 4.3
Age (Mean ± SD)	67.4 ± 7.6	65.8 ± 7.4	59.1 ± 3.9
Age group, n (%) <sup>a</sup>			
55–59 years	324 (16.1)	539 (24.6)	369 (62.8)
60–64 years	540 (26.8)	504 (23.1)	162 (27.5)
65–69 years	436 (21.7)	466 (21.3)	47 (8.0)
70–74 years	308 (15.3)	386 (17.7)	6 (1.0)
75–79 years	231 (11.5)	200 (9.1)	3 (0.5)
80–84 years	138 (6.9)	63 (2.9)	1 (0.2)
≥85 years	35 (1.7)	29 (1.3)	0 (0.0)
Sex, n (%)			
Female	1186 (58.9)	1176 (53.8)	293 (49.8)
Education, n (%)			
Below high school	772 (38.4)	107 (4.9)	5 (0.9)
High school	644 (32.0)	732 (33.5)	100 (17.0)
Some college	361 (17.9)	586 (26.8)	162 (27.5)
College and above	235 (11.7)	762 (34.8)	321 (54.6)

Note: We included participants 55 years or older in the present study. See exclusion criteria in Figure 1. n (%), the frequency and percentage in different groups. See Table S2 in supporting information for additional information.

Abbreviations: NP, neuropsychological; SD, standard deviation.

of each of the NP tests by subtracting the Gen 1 specific mean and dividing by the Gen 1 specific SD across the participants of the three generations. We next standardized each Z score to have a mean of 10 and an SD of one across the generations. Of note, the distribution of Z scores remained unchanged after the standardization procedures. We then calculated a composite domain-specific score as the average of the Z scores of the NP tests in the memory (LMi, LMd, PASi, and VRi) and the attention and executive function (SIM, DSF, and DSB) domains.<sup>29,30</sup>

## 2.4.2 | Outcome and predictor variables

We analyzed two types of outcome variables in this present study: (1) the Z scores of the eight NP tests and the two domain-specific scores, and (2) incident dementia. Comparing the baseline NP test scores across the three generations, the main predictor variable was the generation index. In predicting incident dementia, the main predictor was an NP test score or a composite score. Additional predictor variables included sex (self-reported male and female), age groups, and education groups. The two age groups (55–64 and ≥65 years) and two education groups (high school or lower education, and above high school) were used to characterize the eight NP scores in stratum-specific analyses. The four age groups (55–59, 60–64, 65–69, and ≥70 years) and four education groups (less than high school diploma, high

school graduate, some college, and college or above) were used in regression analyses to compare the NP test scores across the three generations and to predict the development of dementia by an NP test score.

## 2.5 | Statistical analyses

Baseline characteristics were reported as mean ± SD or/and median with interquartile values for continuous variables, and frequencies (percentages) for categorical variables. We characterized the individual NP tests (raw scores and standardized Z scores) in stratified analyses by generation, sex, two education groups, and two age groups. We then used linear mixed regression models to compare whether individual NP scores or domain-specific NP scores were significantly different between generations, adjusting for sex (male and female), four age groups, four education groups, and family correlation. In addition, we conducted Cox proportional hazard regression to investigate whether generation modifies the association of NP test scores with the development of incident dementia. We performed Cox regression analyses in each birth cohort and combined data. We included an interaction term (generation index x NP score) in the combined data to investigate the possible modification effect of generation in the association of NP test scores with incident dementia. Participants without follow-up were excluded from Cox regression analyses. We used SAS (version 9.2) for all statistical analyses. An alpha level of  $P < .05$  was considered significant in all analyses.

## 3 | RESULTS

We reported participant characteristics and illustrated the unadjusted baseline NP test scores between categories of each predictor variable. We then compared the NP test scores between generations by stratified analyses and in linear models. Finally, we demonstrated that the improvement of NP test scores was associated with generation-specific dementia risk. For easier comparison and interpretation, we reported characteristics and regression parameters derived from the standardized NP test scores.

### 3.1 | Participant characteristics at baseline NP tests

Gen 1 participants were born around 1910 (± 6), Gen 2 around 1933 (± 8), and Gen 3 around 1949 (± 4; Table 1). Although we excluded participants younger than 55 years from this study, the average ages were largely different across the three generations: the mean age was 67 years for Gen 1, 66 years for Gen 2, and 59 years for Gen 3 at their baseline NP assessment. With the distribution of 5-year categories across the three generations, Gen 3 included a much higher proportion of younger participants than Gen1 and Gen 2. For example, 62.8% of Gen 3 participants were between 55 and 59 years old, while only 16.1%

of Gen 1 and 24.6% of Gen 2 participants were in this age interval. In contrast, >20% of Gen 1 participants were 75 years or older, while about 13% of participants in Gen 2, and <1% of participants in Gen 2 were 75 years or older (Table 1). Similarly, the educational levels were greatly improved from the older generation to the younger generation at the baseline NP test. The majority (70%) of Gen 1 reported having a high school diploma or lower level of education, and only 11.7% of Gen 1 had a college degree or a higher degree. In contrast, about 38% of Gen 2 and 18% of Gen 3 reported having a high school diploma or lower education level, while about 35% of Gen 2 and 55% of Gen 3 reported having a college or above college degree (Table 2). Gen 1 (58.9% women vs. 41.1% men) and Gen 2 (53.8% women vs. 46.2% men) contained more women than men, while Gen 3 contained equal proportions of men and women (49.8% women vs. 50.2% men).

### 3.2 | Unadjusted baseline NP test measures among generations, sex, age groups, and education groups

The eight NP tests displayed the lowest average levels in Gen 1 and the highest average levels in Gen 3; all pairwise comparisons between generations were significant ( $P < .005$ ) except for the average levels of VRI between Gen 2 and Gen 3 ( $P = .08$ ; Table S3 in supporting information). For example, on average, Gen 2 had 1 SD improvement compared to Gen 1 (10.38 vs. 9.38,  $P < .001$ ) and Gen 3 (score = 10.68) had a 0.3 SD improvement compared to Gen 2 (10.68 vs. 10.38,  $P < .001$ ). Younger participants (<65 years) performed significantly better than older participants ( $\geq 65$  years;  $P < .001$ ) across the three birth cohorts at baseline. On average, younger participants had 0.30 to 0.60 SD higher scores than older participants across the eight NP tests ( $P < .001$ ). More years in school improved all of the eight NP test scores. Participants with college and above degrees displayed 0.37 to 0.60 SD higher scores than those with high school or lower educational levels across the eight NP tests ( $P < .001$ ). No consistent effects of sex on the test scores were found. Without adjustment, the standardized scores of LMi, LMd, and DSB had no significant difference between men and women. In contrast, men appeared to perform slightly better in standardized VRI, SIM, and DSF (0.11 to 0.14 SD higher score than women,  $P < .001$ ). Women performed better in standardized PASi (0.36 SD higher score than men,  $P < .001$ ) and FAS (0.10 SD higher score than men,  $P = .006$ ; Table S3).

### 3.3 | Comparison of baseline NP test scores among generations adjusting for sex, age group, and education group

We characterized the statistical parameters (i.e., number of observations, mean and SD, median, and Q1–Q3) of the eight NP tests (Table 2; Tables S4 and S5 in supporting information). On average, most of the eight NP test scores were lower in Gen 1 than in Gen 2 or Gen 3. This trend was more apparent in older participants ( $\geq 65$  years) compared to the younger participants (55–64 years). For example, the average

standardized LMi scores displayed a 0.83 to 0.96 SD improvement from Gen 1 to Gen 2 and a 0.73 to 1.27 SD improvement from Gen 1 to Gen 3 across the sex and education strata in older participants. The improvement was smaller in younger participants: a 0.55 to 0.77 SD improvement was observed from Gen 1 to Gen 2, and a 0.85 to 0.92 SD improvement from Gen 1 to Gen 3 across the sex and education strata. We observed improvement in LMi, LMd, and PASi test scores but not in other NP tests from Gen 2 to Gen 3 across sex, education, and age strata (Table 2, Figure 2; Tables S4 and S5, Figures S1 and S2 in supporting information).

We then used linear mixed models to quantify the average difference in the standardized NP test scores among birth cohorts adjusting for sex, four education levels, four age groups, and family correlation. Similar to what we observed by stratified analyses, the performance was significantly improved for seven of the eight NP tests in Gen 2 or Gen 3 compared to Gen 1 (a 0.14–0.81 SD improvement,  $P < .001$ ) adjusting for sex, education, and age groups (Table 3). Gen 2 appeared to have a 0.3 SD better performance for VRI in the memory domain ( $P < .0001$ ) and a 0.14 SD better performance for SIM in the attention and executive function domain ( $P = .006$ ) than Gen 3. No significant difference was observed among generations for FAS (Table 3).

### 3.4 | Comparison of domain-specific scores at baseline among generations adjusting for sex, age group, and education group

As expected, similar findings were observed for the domain-specific scores in the memory and attention and executive function domains (Tables 2 and 3). Adjusting for sex, education group, and age groups, we observed better performance in Gen 2 (0.59 SD improvement,  $P < .0001$ ) and Gen 3 (0.52 SD improvement,  $P < .0001$ ) compared to Gen 1 for the composite memory score. Gen 2 performed slightly better than Gen 3 (10.45 vs. 10.37) for the composite memory score although this difference is insignificant ( $P = .35$ ). Similar findings were also observed for the composite attention and executive function domain score among the three generations.

### 3.5 | NP test scores and generation-specific dementia risk

After excluding participants without follow-up, we analyzed 3703 participants (median follow-up 12 years) of Gen 1 ( $n = 1888$ , 650 incident dementia) and Gen 2 ( $n = 1815$ , 209 incident dementia) but not in Gen 3 because Gen 3 had only one incident dementia during follow-up. As expected, better performance was associated with lower hazards of being diagnosed with incident dementia for most of the NP test scores in the combined samples and Gen 1-only and Gen 2-only samples (Table S6 in supporting information; Figure 3). Of note, the insignificant associations among FAS, DSF, and DSB in Gen 2 were likely due to a smaller sample size ( $n = 374$  of 392 Gen 2 in follow-up, 43 incident dementia; Table S2).

**TABLE 2** Statistical parameters of the eight crude neuropsychological test measures stratified by generation, sex, education.

NP test and composite variable	Gen 1				Gen 2				Gen 3			
	High school or lower		Above high school		High school or lower		Above high school		High school or lower		Above high school	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
	Measured NP test, mean (SD)											
	55- to 64- year-old age group											
LMI	7.3(3.3)	6.8(3.2)	9.3(3.6)	9.6(3.1)	9.6(3.4)	10.7(3.2)	11.8(3.2)	12.8(3.1)	10.85(3.4)	11.84(3.4)	11.7(3.1)	13.4(3.3)
LMD	6.4(3.3)	5.5(3.3)	8.4(3.5)	8.0(3.0)	8.7(3.4)	9.8(3.4)	10.9(3.2)	11.9(3.4)	9.57(3.6)	10.56(3.6)	10.5(3.2)	12.5(3.6)
VRI	6.7(3.0)	6.6(2.9)	8.4(3.2)	7.7(3.1)	8.1(3.4)	8.5(2.9)	10.0(2.9)	9.7(2.9)	7.72(2.9)	8.19(2.3)	8.8(2.4)	8.9(2.6)
PASI	12.1(3.2)	13.0(3.5)	13.1(3.1)	14.8(3.1)	12.3(3.0)	14.1(3.0)	13.5(3.2)	15.3(3.1)	13.00(3.0)	14.50(3.6)	13.8(3.4)	15.7(3.3)
SIM	12.2(4.7)	12.1(5.4)	16.2(4.4)	17.1(4.0)	15.2(4.1)	15.5(3.3)	18.0(3.1)	18.0(3.1)	15.86(3.5)	15.23(2.7)	17.7(3.3)	17.6(2.8)
DSF	6.0(1.1)	5.9(1.2)	6.4(1.3)	6.4(1.2)	6.3(1.6)	6.5(1.4)	7.3(1.2)	7.0(1.3)	6.56(1.3)	6.35(1.2)	7.0(1.3)	7.0(1.2)
DSB	4.2(1.3)	4.1(1.1)	4.8(1.5)	5.0(1.4)	4.6(1.6)	4.4(1.1)	5.3(1.4)	5.2(1.2)	4.70(1.3)	4.77(1.1)	5.2(1.4)	5.2(1.3)
FAS	29.5(11.0)	32.5(12.2)	37.9(11.7)	40.7(11.2)	29.2(14.5)	40.6(13.7)	43.4(12.3)	46.84(11.4)	34.76(8.8)	36.65(9.8)	40.0(11.2)	45.0(11.7)
	65±- year-old age group											
LMI	6.0(3.3)	5.7(3.1)	7.8(3.4)	7.6(3.4)	9.4(3.4)	10.1(3.5)	10.8(3.6)	11.6(3.6)	9.0(3.6)	11.2(3.0)	13.0(3.9)	12.8(3.7)
LMD	4.8(3.6)	4.2(3.0)	6.5(3.5)	6.0(3.5)	8.4(3.6)	9.0(3.8)	9.7(3.6)	10.6(3.8)	8.3(4.9)	10.2(3.0)	12.0(4.5)	11.7(4.3)
VRI	4.7(2.9)	4.6(2.7)	6.2(3.1)	6.4(3.0)	7.1(3.2)	6.4(3.1)	8.2(2.9)	8.0(3.2)	7.0(1.7)	6.9(2.1)	8.0(2.2)	8.2(2.3)
PASI	10.7(3.3)	11.8(3.4)	11.5(3.4)	12.9(3.4)	11.6(3.1)	12.8(3.0)	12.3(3.2)	13.9(3.3)	11.5(2.2)	14.6(2.7)	14.1(3.1)	13.8(3.7)
SIM	9.8(5.8)	8.9(5.7)	14.3(5.3)	14.7(4.4)	14.1(4.0)	13.9(3.4)	16.8(3.6)	16.4(3.7)	16.3(2.5)	12.9(4.2)	17.4(3.2)	15.9(3.9)
DSF	5.6(1.3)	5.5(1.3)	6.2(1.2)	6.1(1.2)	6.4(1.3)	6.2(1.3)	6.9(1.3)	6.5(1.3)	5.7(1.2)	6.7(1.2)	7.1(1.3)	6.0(1.1)
DSB	3.8(1.2)	3.9(1.2)	4.5(1.4)	4.4(1.2)	4.6(1.4)	4.4(1.0)	5.0(1.3)	4.9(1.3)	4.0(1.0)	4.8(1.3)	4.8(1.2)	4.8(1.1)
FAS	26.6(11.6)	27.1(12.1)	37.1(12.3)	37.1(11.5)	31.2(12.3)	29.6(10.7)	37.7(12.2)	37.6(14.7)	40.0(8.9)	37.1(14.6)	41.7(13.8)	37.1(10.4)

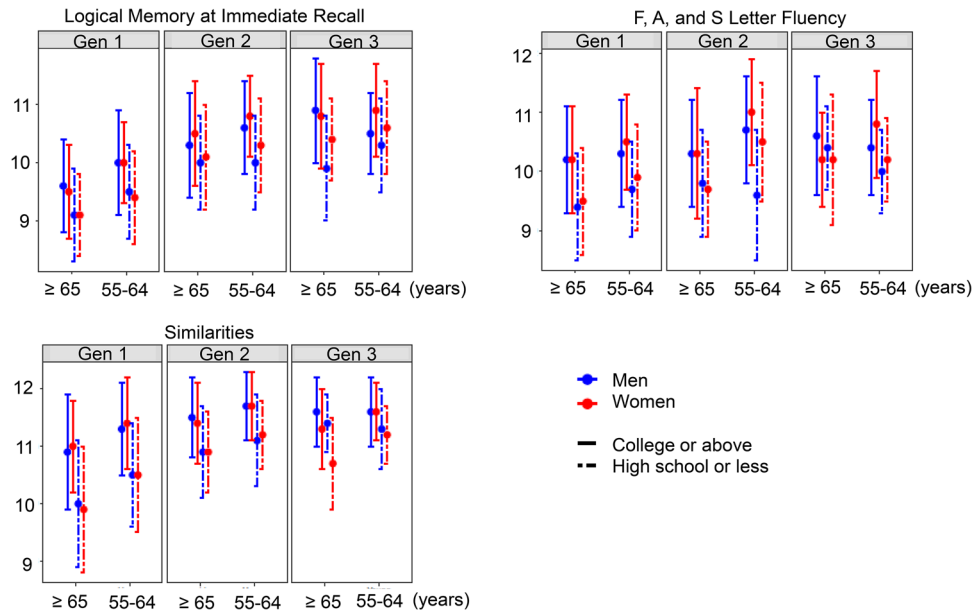
(Continues)

TABLE 2 (Continued)

NP test and composite variable	Gen 1				Gen 2				Gen 3			
	High school or lower		Above high school		High school or lower		Above high school		High school or lower		Above high school	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Standardized NP score, mean (SD)												
<u>55- to 64- year-old age group</u>												
LMI	9.5 (0.8)	9.4 (0.8)	10.0 (0.9)	10.0 (0.7)	10.0 (0.8)	10.3 (0.8)	10.6 (0.8)	10.8 (0.7)	10.3 (0.8)	10.6 (0.8)	10.5 (0.7)	11.0 (0.8)
LMd	9.5 (0.8)	9.3 (0.8)	10.0 (0.8)	9.9 (0.7)	10.1 (0.8)	10.3 (0.8)	10.6 (0.7)	10.8 (0.8)	10.3 (0.8)	10.5 (0.8)	10.6 (0.8)	11.0 (0.8)
VRi	9.8 (0.9)	9.8 (0.8)	10.3 (1.0)	10.1 (0.9)	10.2 (1.0)	10.3 (0.9)	10.8 (0.9)	10.7 (0.9)	10.1 (0.9)	10.2 (0.7)	10.4 (0.7)	10.5 (0.8)
PASi	9.7 (0.9)	10.0 (1.0)	10.0 (0.9)	10.5 (0.9)	9.8 (0.9)	10.3 (0.9)	10.1 (0.9)	10.7 (0.9)	10.0 (0.9)	10.4 (1.0)	10.2 (1.0)	10.8 (0.9)
SIM	9.5 (0.9)	9.5 (1.0)	10.3 (0.8)	10.4 (0.8)	10.1 (0.8)	10.2 (0.6)	10.7 (0.6)	10.7 (0.6)	10.3 (0.7)	10.2 (0.5)	10.6 (0.6)	10.6 (0.5)
DSF	9.9 (0.8)	9.8 (0.9)	10.2 (0.9)	10.2 (0.9)	10.1 (1.2)	10.2 (1.0)	10.8 (0.9)	10.6 (1.0)	10.3 (0.9)	10.1 (0.8)	10.6 (1.0)	10.6 (0.9)
DSB	9.8 (0.9)	9.8 (0.8)	10.3 (1.1)	10.4 (1.0)	10.2 (1.2)	10.0 (0.8)	10.7 (1.0)	10.5 (0.9)	10.2 (1.0)	10.3 (0.8)	10.6 (1.0)	10.6 (1.0)
FAS	9.7 (0.8)	9.9 (0.9)	10.3 (0.9)	10.5 (0.8)	9.6 (1.1)	10.5 (1.0)	10.7 (0.9)	11.0 (0.9)	10.1 (0.7)	10.2 (0.7)	10.4 (0.8)	10.8 (0.9)
Memory	9.5 (0.8)	9.5 (0.8)	10.1 (0.9)	10.2 (0.8)	10.2 (0.8)	10.5 (0.8)	10.5 (0.7)	11.0 (0.8)	10.0 (0.8)	10.4 (0.8)	10.6 (0.7)	10.9 (0.8)
Attention	9.8 (0.8)	9.7 (0.8)	10.4 (0.8)	10.5 (0.8)	10.4 (0.7)	10.3 (0.6)	10.9 (0.8)	10.9 (0.7)	10.1 (1.1)	10.3 (0.7)	11.0 (0.7)	10.9 (0.7)
<u>65± -year-old age group</u>												
LMI	9.2 (0.8)	9.1 (0.7)	9.6 (0.8)	9.5 (0.8)	10.0 (0.8)	10.2 (0.9)	10.3 (0.9)	10.5 (0.9)	9.88 (0.9)	10.4 (0.7)	10.9 (0.9)	10.8 (0.9)
LMd	9.2 (0.8)	9.0 (0.7)	9.6 (0.8)	9.4 (0.8)	10.0 (0.8)	10.2 (0.9)	10.3 (0.8)	10.5 (0.9)	9.99 (1.1)	10.4 (0.7)	10.9 (1.0)	10.8 (1.0)
VRi	9.2 (0.8)	9.2 (0.8)	9.6 (0.9)	9.7 (0.9)	9.9 (1.0)	9.7 (0.9)	10.3 (0.9)	10.2 (0.9)	9.88 (0.5)	9.9 (0.6)	10.2 (0.7)	10.2 (0.7)
PASi	9.3 (0.9)	9.6 (1.0)	9.6 (1.0)	10.0 (1.0)	9.6 (0.9)	9.9 (0.9)	9.8 (0.9)	10.2 (0.9)	9.56 (0.6)	10.5 (0.8)	10.3 (0.9)	10.2 (1.1)
SIM	9.0 (1.1)	8.9 (1.1)	9.9 (1.0)	10.0 (0.8)	9.9 (0.8)	9.9 (0.7)	10.5 (0.7)	10.4 (0.7)	10.36 (0.5)	9.7 (0.8)	10.6 (0.6)	10.3 (0.7)
DSF	9.6 (1.0)	9.5 (0.9)	10.1 (0.9)	10.0 (0.9)	10.2 (0.9)	10.0 (0.9)	10.6 (1.0)	10.3 (1.0)	9.63 (0.9)	10.4 (0.9)	10.7 (0.9)	9.9 (0.8)
DSB	9.5 (0.9)	9.6 (0.9)	10.0 (1.0)	10.0 (0.9)	10.1 (1.1)	9.9 (0.7)	10.4 (1.0)	10.3 (1.0)	9.68 (0.7)	10.3 (0.9)	10.3 (0.9)	10.3 (0.8)
FAS	9.4 (0.9)	9.5 (0.9)	10.2 (0.9)	10.2 (0.9)	9.8 (0.9)	9.7 (0.8)	10.3 (0.9)	10.3 (1.1)	10.44 (0.7)	10.2 (1.1)	10.6 (1.0)	10.2 (0.8)
Memory	9.0 (0.8)	9.0 (0.8)	9.5 (0.8)	9.6 (0.8)	9.9 (0.8)	10.0 (0.8)	10.2 (0.8)	10.5 (0.8)	9.77 (1.0)	10.4 (0.6)	10.7 (0.9)	10.6 (1.0)
Attention	9.3 (0.9)	9.3 (0.9)	10.0 (0.8)	10.1 (0.8)	10.2 (0.9)	10.0 (0.7)	10.7 (0.8)	10.4 (0.9)	9.97 (0.5)	10.3 (0.8)	10.8 (0.7)	10.3 (0.6)

Notes: A subset of Gen 2 (n = 392) participants were measured for the three tests (F, A, and S Letter Fluency, Digit Span Forward, and Digit Span Backward) at the baseline NP test. Memory, the memory function domain was the average scores of LMI, LMd, PASi, and VRi. Attention, the attention function domain was calculated as the average of SIM, DSB, and DSB. FAS is only one measure in the verbal fluency domain. See Methods for details.

Abbreviations: DSB, Digit Span Backward; DSF, Digit Span Forward; FAS, F, A, and S Letter Fluency; LMI, Logical Memory at Delayed Recall; LMIi, Logical Memory at Immediate Recall; NP, neuropsychological; PASi, Paired Associate Learning (Immediate Recall); SD, standard deviation; SIM, Similarities; VRi, Visual Reproduction at Immediate Recall.



**FIGURE 2** Characterization of the standardized NP test scores among generations in each sex, education group, and age group. We displayed the mean and standard deviation of standardized NP test scores of Logical Memory at Immediate Recall in the memory function domain, Similarities in the attention and executive function domain, and of F, A, and S Letter Fluency in the verbal fluency function domain. NP, neuropsychological.

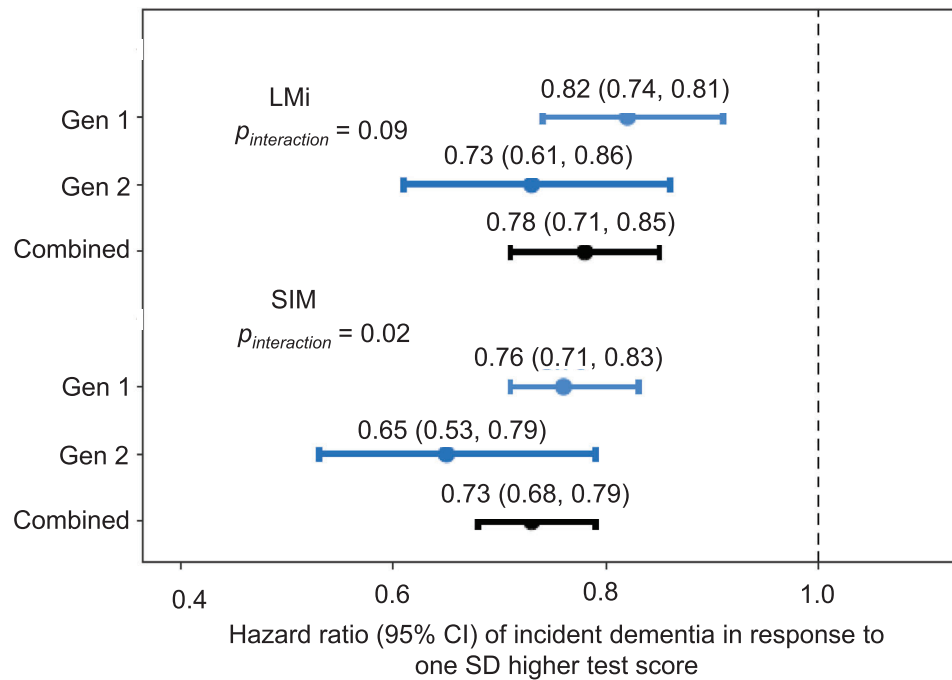
**TABLE 3** Comparison of the standardized NP scores and composite scores among generation cohorts.

NP test or composite score	Gen 1 (n = 1012)	Gen 2 (n = 2187)	Gen 3 (n = 588)	P value (Gen 1 vs. Gen 2)	P value (Gen 1 vs. Gen 3)	P value (Gen 2 vs. Gen 3)
Individual tests						
LMi	9.53 (0.019)	10.25 (0.018)	10.33 (0.036)	<.0001	<.0001	.1
LMd	9.51 (0.019)	10.29 (0.018)	10.32 (0.036)	<.0001	<.0001	.56
VRi	9.75 (0.020)	10.20 (0.020)	9.9 (0.039)	<.0001	.0013	<.0001
PASi	9.86 (0.022)	10.00 (0.021)	10.10 (0.043)	<.0001	<.0001	.062
SIM	9.71 (0.018)	10.14 (0.018)	10.00 (0.035)	<.0001	<.0001	.006
DSF	9.92 (0.023)	10.30 (0.048)	10.24 (0.044)	<.0001	<.0001	.56
DSB	9.95 (0.022)	10.23 (0.048)	10.20 (0.044)	<.0001	<.0001	.86
FAS	9.99 (0.021)	10.09 (0.046)	10.11 (0.042)	.10	.028	.93
Composite scores						
Memory	9.72 (0.022)	10.31 (0.022)	10.29 (0.035)	<.0001	<.0001	.82
Attention	10.01 (0.014)	10.45 (0.054)	10.37 (0.037)	<.0001	<.0001	.35

*Notes:* This study included participants who were aged 55 years and older at baseline NP test. We compared the NP standardized scores and domain-specific scores among generations, adjusting for age groups, sex, and education levels in a linear mixed model (see Methods). The results were displayed as least-square means with standard error in linear mixed model. A subset of Gen 2 ( $n = 392$ ) participants were measured for the three tests (Digit Span Forward, Digit Span Backward, and F, A, and S Letter Fluency) at the baseline NP test. Memory, the memory function domain was the average scores of LMi, LMd, PASi, and VRi. Attention, the attention function domain, was calculated as the average of SIM, DSB, and DSB. FAS is only one measure in the verbal fluency domain. See Methods for details.

Abbreviations: DSB, Digit Span Backward; DSF, Digit Span Forward; FAS, F, A, and S Letter Fluency; LMd, Logical Memory at Delayed Recall; LMi, Logical Memory at Immediate Recall; NP, neuropsychological; PASi, Paired Associate Learning (Immediate Recall); SD, standard deviation; SIM, Similarities; VRi, Visual Reproduction at Immediate Recall.





**FIGURE 3** Generation-specific dementia hazard with standardized scores in Logical Memory at Immediate Recall and Similarities. We performed association analyses of the development of incident dementia with the standardized scores in NP tests in the Framingham Heart Study Original cohort (Gen 1), Offspring cohort (Gen 2). We observed a significant modification effect of generation on the association of Similarities (SIM) with incident dementia hazard ( $P_{interaction} = .02$ ) while the modification of generation on Logical Memory at Immediate Recall (LMI) with incident dementia hazard was insignificant ( $P_{interaction} = .09$ ). See Table S6 in supporting information for additional results of other NP test scores. NP, neuropsychological.

In cohort-specific analyses, we found that generation displayed effect modification on the association between the NP test scores and incident dementia. We observed that better performance in the NP test scores was associated with larger reductions of being diagnosed with incident dementia in Gen 2 than in Gen 1 (Table S6). However, the effect modification was significant only for the SIM test with incident dementia. One SD improvement in SIM performance was associated with a 35% reduction (hazard ratio [HR] = 0.65, 95% confidence interval [CI] = 0.53, 0.79) in dementia risk in Gen 2 and with a 24% reduction (HR = 0.76, 95% CI = 0.71, 0.83) in dementia risk in Gen 1 (35% vs. 24%,  $P_{interaction} = .02$  in the combined Gen 1 and Gen 2 samples; Figure 3; Table S6).

## 4 | DISCUSSION

In this study, we compared the baseline cognitive performance in late middle-aged and older participants among three FHS generations to better understand the effects of birth cohorts on cognitive performance. We observed that Gen 2 and Gen 3 performed significantly better on seven of eight cognitive tests than Gen 1 on the baseline NP tests. We also observed that the younger cohort tended to have larger reductions in being diagnosed with incident dementia per SD improvement in the cognitive tests. To our knowledge, this is the first attempt to investigate the differences in cognitive performance with standard NP

tests across multiple generations of the same population over several decades.

In FHS, the baseline cognitive assessment was conducted for three generations, respectively, in the late 1970s, around the millennium, and in 2010. These time points represent the three major eras for looking back at how changes in education, technology, lifestyle, and health care have impacted our society, creating a more cognitively stimulating environment and resulting in the rapid increase in cognitive test scores that were observed in other parts of world.<sup>9,31,32</sup> As expected, we found that later generations performed significantly better than the first generation. In addition, an SD improvement in the cognitive tests was associated with larger reductions of being diagnosed with incident dementia in the FHS Offspring cohort compared to their parents. These findings have provided complementary information to the observations that the dementia incidence rate has been declining in recent years.<sup>1-4</sup> Our findings demonstrate the need for the continuing evaluation of normative data due to birth cohort effects that are tied to evolving changes in many areas of our society.

Age structure and education levels differed considerably across generations in FHS, and thus, age and education were accounted for in comparing NP test scores among the generations. However, this method may bias the association toward the null because the generation index variable was highly correlated with age and education. Nonetheless, we still observed significantly better performance by Gen 2 and Gen 3 than Gen 1 for most of the NP test scores. The

present study only compared cognitive tests cross-sectionally at baseline, which does not reflect the secular changes in cognitive function in the same birth cohort. Several important factors, including socioeconomic status, dietary factors, health variables, and other biological markers, were not investigated in this study. In addition, this study only centered on late middle-aged and older adults. For this reason, only a small subset of Gen 3 participants was included in the study. Thus, the findings in Gen 3 may not represent the entire Gen 3 cohort in generation comparative analyses. Future studies are warranted to investigate the NP normative values in younger adults across generations for the increasing focus on dementia/Alzheimer's disease as a life-course disease.<sup>33</sup> The reduction in sample size for DSF, DSB, and FAS in Gen 2 may result in bias when comparing the three NP tests in association analyses. This study only included non-Hispanic White individuals with similar social backgrounds. Therefore, our findings may not be easily generalizable to other US populations. Despite multiple weaknesses in this study, our findings have provided complementary information, at least, to explain partially the observed decline in the dementia incidence rate in recent years.<sup>1-4</sup>

In conclusion, the findings have illustrated the need to recognize that, in cognitive assessment, the normative scores to determine significant cognitive changes should be tailored to each generation. Thus, generation-based or cohort-based norms should be applied when identifying potential risk factors and insight for the diagnosis of cognitive impairment and dementia at the present time.

#### ACKNOWLEDGMENT

We would like to thank the Framingham participants for their support. Data collection for FHS was supported by N01-HC-25195 and HHSN268201500001. Other funding resources include R01AG059727, R01AG016495, R01AG008122, R01AG062109, and U19 AG068753 from the National Institute on Aging.

#### CONFLICT OF INTEREST STATEMENT

RA serves on scientific advisory boards for Biogen and Signant Health.

#### CONSENT STATEMENT

All participants provided written informed consent. The institutional review board at Boston University School of Medicine approved the study procedures.

#### REFERENCES

- Rocca WA, Petersen RC, Knopman DS, et al. Trends in the incidence and prevalence of Alzheimer's disease, dementia, and cognitive impairment in the United States. *Alzheimers Dement*. 2011;7(1):80-93.
- Winblad B, Amouyel P, Andrieu S, et al. Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet Neurol*. 2016;15:455.
- Wu YT, Beiser AS, Breteler M, Fratiglioni L, Brayne C. The changing prevalence and incidence of dementia over time — current evidence. 2017.
- Satizabal CL, Beiser AS, Chouraki V, Chêne G, Seshadri S. Incidence of dementia over three decades in the Framingham Heart Study. *N Engl J Med* 2016;375:523-532.
- Petersen RC. *Mild cognitive impairment : aging to Alzheimer's disease*. Oxford University Press, 2003.
- Harvey PD. Clinical applications of neuropsychological assessment. *Dialogues Clin Neurosci* 2012;14:91-99.
- Flynn JR. The mean IQ of Americans: massive gains 1932 to 1978. *Psychol Bull* 1984;95:29-51.
- Flynn JR. Searching for justice: the discovery of IQ gains over time. *Am Psychol* 1999;54:5-20.
- Salthouse TA. When does age-related cognitive decline begin? *Neurobiol Aging* 2009;30:507-514.
- Ang S, Rodgers JL, Wanstrom L. The Flynn effect within subgroups in the U.S.: gender, race, income, education, and urbanization differences in the NLSY-children data. *Intelligence* 2010;38:367-384.
- Platt JM, Keyes KM, McLaughlin KA, Kaufman AS. The Flynn effect for fluid IQ may not generalize to all ages or ability levels: a population-based study of 10,000 US adolescents. *Intelligence* 2019;77.
- Farmer ME, White LR, Kittner SJ, et al. Neuropsychological test performance in Framingham: a descriptive study. *Psychol Rep* 1987;60:1023-1040.
- Elias MF, Elias PK, D'Agostino RB, Silbershatz H, Wolf PA. Role of age, education, and gender on cognitive performance in the Framingham Heart Study: community-based norms. *Exp Aging Res* 1997;23:201-235.
- Elias MF, Elias PK, Sullivan LM, Wolf PA, D'Agostino RB. Obesity, diabetes and cognitive deficit: the Framingham Heart Study. *Neurobiol Aging* 2005;26 Suppl 1:11-16.
- Au R, Seshadri S, Wolf PA, et al. New norms for a new generation: cognitive performance in the Framingham offspring cohort. *Exp Aging Res* 2004;30:333-358.
- DeBette S, Beiser A, DeCarli C, et al. Association of MRI markers of vascular brain injury with incident stroke, mild cognitive impairment, dementia, and mortality: the Framingham Offspring Study. *Stroke* 2010;41:600-606.
- Murabito JM, Beiser AS, DeCarli C, Seshadri S, Wolf PA, Au R. Parental longevity is associated with cognition and brain ageing in middle-aged offspring. *Age Ageing* 2014;43:358-363.
- Pase MP, Himali JJ, Mitchell GF, et al. Association of aortic stiffness with cognition and brain aging in young and middle-aged adults: the Framingham third generation cohort study. *Hypertension* 2016;67:513-519.
- Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* 1951;41:279-281.
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med* 1975;4:518-525.
- Splansky GL, Corey D, Yang Q, et al. The third generation cohort of the national heart, lung, and blood institute's Framingham heart study: design, recruitment, and initial examination. *Am J Epidemiol* 2007;165:1328-1335.
- Sacco RL, Wolf PA, Kannel WB, McNamara PM. Survival and recurrence following stroke. The Framingham study. *Stroke* 1982;13:290-295.
- Bachman DL, Wolf PA, Linn R, et al. Prevalence of dementia and probable senile dementia of the Alzheimer type in the Framingham Study. *Neurology* 1992;42:115-119.
- Wechsler D, Wechsler D. The Wechsler-Bellevue intelligence scale. *The Psychological corporation*, 1946.
- Wechsler DJe. Book Reviews: The Measurement and Appraisal of Adult Intelligence. 1958;128:1133.
- Benton AL. *Multilingual aphasia examination*. University of Iowa Press. 1976.
- Au R, Piers RJ, Devine S. How technology is reshaping cognitive assessment: Lessons from the Framingham Heart Study. *Neuropsychology* 2017;31:846-861.

28. Mckhann G, Drachman D, Folstein M, Katzman R, Price D. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984.
29. Chapman RM, Mapstone M, Porsteinsson AP, et al. Diagnosis of Alzheimer's disease using neuropsychological testing improved by multivariate analyses. *J Clin Exp Neuropsychol* 2010;32: 793-808.
30. Downer B, Fardo DW, Schmitt FA. A summary score for the Framingham heart study neuropsychological battery. *J Aging Health* 2015;27:1199-1222.
31. Flynn, James RJPB. The mean IQ of Americans: massive gains 1932 to 1978. *Psychol Bull* 1984;95:29-51.
32. Flynn J. Searching for justice: the discovery of IQ gains over time. *Am Psychol* 1999;54:5-20.
33. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. *Lancet* 2017;390:2673-2734.

### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Yang J, Ang TFA, Lu S, et al. Establishing cognitive baseline in three generations: Framingham Heart Study. *Alzheimer's Dement*. 2023;15:e12416. <https://doi.org/10.1002/dad2.12416>