Supplementary Information for

Influence of Young Adult Cognitive Ability and Additional Education on Later Life Cognition

William S. Kremen, Asad Beck, Jeremy A. Elman, Daniel E. Gustavson, Chandra A. Reynolds, Xin M. Tu, Mark E. Sanderson-Cimino, Matthew S. Panizzon, Eero Vuoksimaa, Rosemary Toomey, Christine Fennema-Notestine, Donald J. Hagler, Jr., Bin Fang, Anders M. Dale, Michael J. Lyons, Carol E. Franz

Corresponding author: William S. Kremen
Email: wkremen@ucsd.edu

This PDF file includes**:**

Supplementary text
Fig. S1
Tables S1 to S8
References for SI reference citations

Supplementary Information Text

Materials and Methods

Measures. Childhood socioeconomic status (cSES), education, occupation, and health status (based on report of diagnosed medical conditions) were determined from structured interviews with study participants. Engagement in cognitive-intellectual activities and physical activity were questionnaire-based. As indicated in the main text, occupational complexity, engagement in cognitive-intellectual activities, and physical activity were based on the VETSA 1 assessment when participants were 51-60 years old. Occupational complexity was based on the highest occupation that the participants had attained up to the time of assessment. Occupational classifications were based on the International Standard Classification of Occupations (ISCO) (1). The assigned numbers (see *Results*) are not integers. Rather, each digit refers to specific occupation coding (e.g., major group, minor group, etc.). At average age 56, it is quite rare for someone to have had a very recent shift to a substantially more complex type of occupation. Engagement in cognitive-intellectual activities and physical activity were based on the past month at the time of the VETSA 1 assessment. Correlations with the same measures at VETSA

2 suggest that the 1-month timeframe provides reasonably stable estimates. Spearman rank order correlations over this approximately 6-year interval were ρ=0.644 (p=1.867e-118) for engagement in cognitive activities and ρ=0.545 (p=1.196e-78) for physical activity.

Cognitive Measures. These were assessed face-to-face. Seven cognitive domains were derived from 23 scores from 13 neuropsychological tests administered during VETSA 2 at average age 62 (2,3). Most of these are widely used tests in clinical neuropsychological assessment. Although all neuropsychological tests are multi-determined, these tests are generally organized into the cognitive domains described based on what are considered the predominant ability being tapped. These determinations come from over a century of research and clinical studies of the spared and impaired functions of people with various types of brain disease or damage, or neuropsychiatric conditions (4). The results stem from neuropsychological testing and are supported by structural and functional neuroimaging studies and postmortem studies (4). Domain scores were created by first *z*-scoring test scores. As commonly done in neuropsychological studies, for domains with multiple tests and scores, we calculated the mean of *z*-scored measures included within each domain. Only the executive function domain score was calculated differently. Based on several prior studies of executive function, testing comprising this domain was subjected to a factor analysis which generated a common executive factor (5). We then derived factor scores for this executive function factor. We have presented results for these cognitive domains in several prior publications (e.g., 2,3). The cognitive domains were: abstract reasoning; episodic memory; processing speed; verbal fluency; visual-spatial ability; ; short-term/working memory; and executive function.

Age-scaled scores are not available for all of these measures. *Z*-scores were used to create the cognitive domain scores from the raw scores. Age and race/ethnicity were then used as covariates in all of the models. We adjusted cognitive domain scores for age and race/ethnicity prior to entering them into the analyses. In other words, the analyses included residualized cognitive domain scores after adjusting for age and race/ethnicity.

Table S3 shows the specific tests and scores comprising each cognitive domain.

General Cognitive Ability (GCA): Armed Forces Qualification Test (AFQT). The same version of this test was administered at average age 20 and again at VETSA 2 (average age 62). The AFQT provides a well-validated GCA measure (6-8). The AFQT is a paper-and-pencil, multiple-choice test with items covering vocabulary, arithmetic word problems, visual-spatial processing, and reasoning about tools and mechanical relations. In VETSA and in other studies, the AFQT correlates about 0.85 with Wechsler IQ, and it was correlated 0.73 across a 4-decade interval in VETSA (6,7).

The cognitive domains are as follows:

Abstract Reasoning: Wechsler Abbreviated Scale of Intelligence Matrix Reasoning subtest (9). This domain includes only a single test. The test comprises trials with a series of designs, and the subject much decide which of 5 choices would be the next design in the sequence.

Episodic Memory: California Verbal Learning Test-II (CVLT) short-delay recall, long-delay recall, total of trials 1-5 (10); Wechsler Memory Scale (WMS-III) Logical Memory, Visual Reproduction subtests (11). The CVLT has 5 learning trials, each consisting of a list of 16 words, followed by an interference list and then short- and long-delay free recall of the repeated list. Logical Memory includes 2 brief stories that are read to the subject. There is an immediate free recall after each story and then a delayed recall. Visual Reproductions involved immediate and delayed recall of 5 designs. Immediate recall follows presentation of each design for 10 seconds.

Processing Speed: Number of words generated on the Stroop word condition, color condition;(12) D-KEFS Trail Making Test number sequencing condition, letter sequencing (13). The Stroop word condition includes the words red, green, and blue in random order written in black ink down columns on a page. In this version, the subject must read the words as fast as possible, and the score is the number of correct words read in 45 seconds. The color condition the same except that instead of words there are 4 colored Xs (XXXX) in the 3 colors, and the

subject must name the colors. In Trails number sequencing, the subject must draw a line connecting circles with numbers in them in order from 1 through 16. The circles are pseudo-randomly placed on the page. The score is the time to complete the task. Letter sequencing is the same except that there are letters in the circles and the subject must go in alphabetical order.

Verbal Fluency: Delis-Kaplan Executive Function System (D-KEFS) Verbal Fluency phonemic (3 trials); semantic (2 trials) fluency (13). There were 3 phonemic trials; saying as many words as possible in 60 seconds that begin with the letters F, A, and S. There were 2 semantic trials in which the subject must say as many words as possible that belong in a particular category: animals and boys' names.

Visual-Spatial Ability: Card Rotations; (14) Hidden Figures (15). Card Rotations is a mental rotation task. Subjects view a design and then must indicate if other rotated designs in that set are the same as the original. In Hidden Figures, subjects are shown a design at the top of the page and then must find that design embedded in other more complex designs.

Working Memory: Reading Span;(16) WMS-III Digit Span, Letter-Number Sequencing, Spatial Span subtests (11). In Reading Span, subjects are presented sentences on a computer monitor. They must read each sentence aloud without pausing between sentences. After a set of sentences is presented, the subject must recall the last word of each sentence. There are sets of 2, 3, and 4 sentences. Letter-Number Sequencing is like Digit Span except that both letters and numbers are presented. Subjects must repeat them but reorganized so that numbers are reported first in order followed by letters in alphabetical order. Spatial Span is visual analog to Digit Span. Subjects must tap series of blocks in the same order as done by the examiner, and in the reverse order in the backwards condition.

Executive Function: Common executive function factor based on Stroop interference condition; DKEFS Trails switching condtion; D-KEFS catergory fluency switching condition (5) Scores were residual scores adjusted for non-interference/non-switching conditions. A similar

factor has been found in other samples as well (17,18). The Stroop interference condition involves saying the color of the ink that a color word is printed in while ignoring the word (e.g., when the word blue is printed in red ink, the subject must say red instead of blue). Trails switching involves circles with numbers and circles with letters on the page. Subjects must connect them in sequence while alternating between number and letter. Category switching involves as many words as possible in 60 seconds from the categories "fruit" and "furniture" while alternating categories.

MRI Acquisition. Images were acquired at two sites, UCSD (n = 256) and MGH (n = 164). At UCSD, images were acquired with a GE 3T Discovery 750× scanner (GE Healthcare, Waukesha, WI, USA) with an 8-channel phased array head coil. The imaging protocol included a sagittal 3D fast spoiled gradient echo (FSPGR) T_1-weighted volume optimized for maximum gray/WM contrast [TE=3.164 msec, TR=8.084 msec, TI=600 msec, flip angle=8°, pixel bandwidth=244.141, matrix=256x192, in-plane resolution=1x1 mm, slice thickness=1.2 mm, slices=172].

At MGH, images were acquired with a Siemens Tim Trio, (Siemens USA, Washington, D.C.) with a 32-channel head coil. The imaging protocol included a 3D magnetization-prepared rapid gradient-echo (MPRAGE) T_1-weighted volume optimized for maximum gray/WM contrast [TE=4.33 msec, TR=2170 msec, TI=1100 msec, flip angle=7°, pixel bandwidth=140, matrix= 256x256, in-plane resolution=1x1 mm, slice thickness=1.2 mm, slices=160].

MRI Processing. The structural MR images were processed as described previously (19-22). We used the FreeSurfer 5.1 (surfer.nmr.mgh.harvard.edu) software package for morphometric analysis of the cortical surface (23-25). Preprocessing included correction of distortion due to gradient nonlinearity (26), image intensity normalization, and rigid registration into standard orientation with 1 mm isotropic voxel size. Boundaries between gray matter, white matter, and cerebral spinal fluid were defined. Based on all vertices comprising the cortical surface, we obtained measures of total surface area and mean cortical thickness for MRIs obtained at

VETSA 2 (average age 62; range: 56.50-66.50). All images required some form of manual intervention to ensure the correct classification of the white matter and pial surfaces, either with normalization control points or manual editing of white matter or brain masks. Particular attention was given to the orbitofrontal cortex, temporal lobes, meninges, and transverse and superior sagittal sinuses. Problematic cortical surface reconstructions were reviewed by consensus with 3 neuroimaging analysts. Ten reconstructions were unable to be corrected and were excluded from this investigation, resulting in a final total of 359 participants. After processing, editing, and quality control, there were 367 individuals with cognitive data and analyzable imaging data.

Multiple Testing Correction. We used the method of Li and Jia variation of the popular the false discovery rate (FDR) developled by Benjamini and Hochberg (27) to more effectively control for multiple comparisons by accounting for correlations among multiple outcomes (dependent variables in regression models).

For independent tests: Consider testing m number of hypotheses. Let α denote the overall (or family-wise) rate for testing the m hypotheses such as $\alpha = 0.05$. Because Bonferroni is generally too conservative, especially for large m (e.g., m > 5), the FDR provides a less stringent alternative. Let p_k denote the ordered p-values for the tests from the smallest to the largest ($1 \le k \le m$). The procedure determines the statistical significance of each test based on the following steps:

a. Compare $p_i \le \frac{i}{m}\alpha$ and find the k that is the largest i for the above to hold true, i.e.,

$$p_k \le \frac{k}{m}\alpha \text{ and } p_{k+1} > \frac{k+1}{m}\alpha.$$

b. Reject the null hypotheses corresponding to the first k smallest p-values.

For dependent tests: In most applications, tests may not be independent. For example, consider m regression models, each with r number of independent variables (excluding the

intercept). If the m outcomes (dependent variables) arecorrelated, the method above is not optimal, because it does not account for such correlations. For example, in the extreme case that the m outcomes are perfectly correlated, there is no adjustment needed for the m regression models and FDR is applied to only the r tests.

Li and Ji discussed an approach to account for correlations among multiple outcomes (28). Their procedure is readily applied to FDR (or Bonferroni) to account for such correlations when used for multiple regression models as in our study. By applying their approach, we first computed the effective number of tests, m_{eff}, based on the multiple outcomes and then used that number to modify the procedures for multiple comparison. When using FDR, we replace the comparison $p_i \le \frac{i}{m}\alpha$ in the FDR procedure in (a) above with the following:

$$p_i \le \frac{\alpha}{m_{eff}} + \frac{i-1}{m-1}\left(\alpha - \frac{\alpha}{m_{eff}}\right).$$

The above works especially well, when m is large and the outcomes are moderately or highly correlated. Using this method, tests with p-values below the following levels were determined to be statistically significant:

1. Model 1 (Table 2): p< 0.022.
2. Model 2 (Table S4): p<0.018.
3. Model 3 (Table 3): p<0.02.
4. Model 4 (Table 4): p<0.009.

Model 5 (Table S5), the analysis with the dichotomized education variable, was performed for the age 62 GCA outcome only for comparison with the study of Clouston et al. (29). Only 2 other predictors were included (cSES and age 20 GCA), so there were not concerns about multiple test correction. Significance levels for the 3 tests were p=0.346, p=8.3e-52, and p=.007.

In Table S6, the analyses with only a single other predictor in addition to age 20 GCA

(either education or occupational complexity), all but 2 tests were highly significant (ps<8.37e-05). The 2 remaining tests were nonsignificant (ps≥0.064).

Cotwin-Control Analysis. Follow-up analyses were conducted using a cotwin-control design to evaluate whether the observed associations among the predictor variables with cognitive performance outcomes may show evidence of a direct causal effect, i.e., after controlling for genetic and familial/shared environmental effects (30,31). The cotwin-control design evaluates within-monozytic (MZ) and within-dizygotic (DZ) pair differences as predictive of within-pair differences in cognitive outcomes, controlling for potential shared genetic or environmental confounders. Thus, if lower education contributes causally to poorer cognitive performance, we expect the MZ twin with lower educational attainment to show poorer cognitive functioning. Equation 1 below represents the within- and between- pair model tested:

$$Cog_{ij} = \beta_0 + \beta_W\left(LifeED_{ij} - \overline{LifeED}_{.j}\right) + \beta_B\left(\overline{LifeED}_{.j}\right) + \varepsilon_{ij} \quad [1]$$

where Cog_{ij} is the outcome for the ith participant within the jth twin pair (j=1,…,N), $LifeED_{ij}$ is their corresponding lifetime education level, and $\overline{LifeED}_{.j}$ is the average lifetime education for the jth twin pair. The residual, ε_{ij}, represents unexplained variation in cognitive performance that is correlated within but not across twin pairs, and β_0 is the intercept. β_B estimates the effect of lifetime education that is shared among the pairs; it should approximate the unadjusted education-cognition association (when unadjusted for any confounders the pairs may share). The within-pair coefficient, β_W, represents the lifetime education-cognition effect adjusted for all shared confounders among members of the twin pairs. If β_W is significant it supports a direct environmental effect. Moreover, if truly environmental in nature we expect the same magnitude of association among DZ pairs, and to the total sample in a typical regression of cognition on the lifetime education predictor. However, if genetic confounding exists the within-pair effect among the DZ pairs will be stronger than for MZ pairs, and the strongest yet would be observed

in the typical regression of cognition on the education predictor. The model in equation 1 was expanded to include interaction terms with zygosity for the within-pair effect term to estimate separate within-pair effects (β_W) for MZ and DZ pairs.

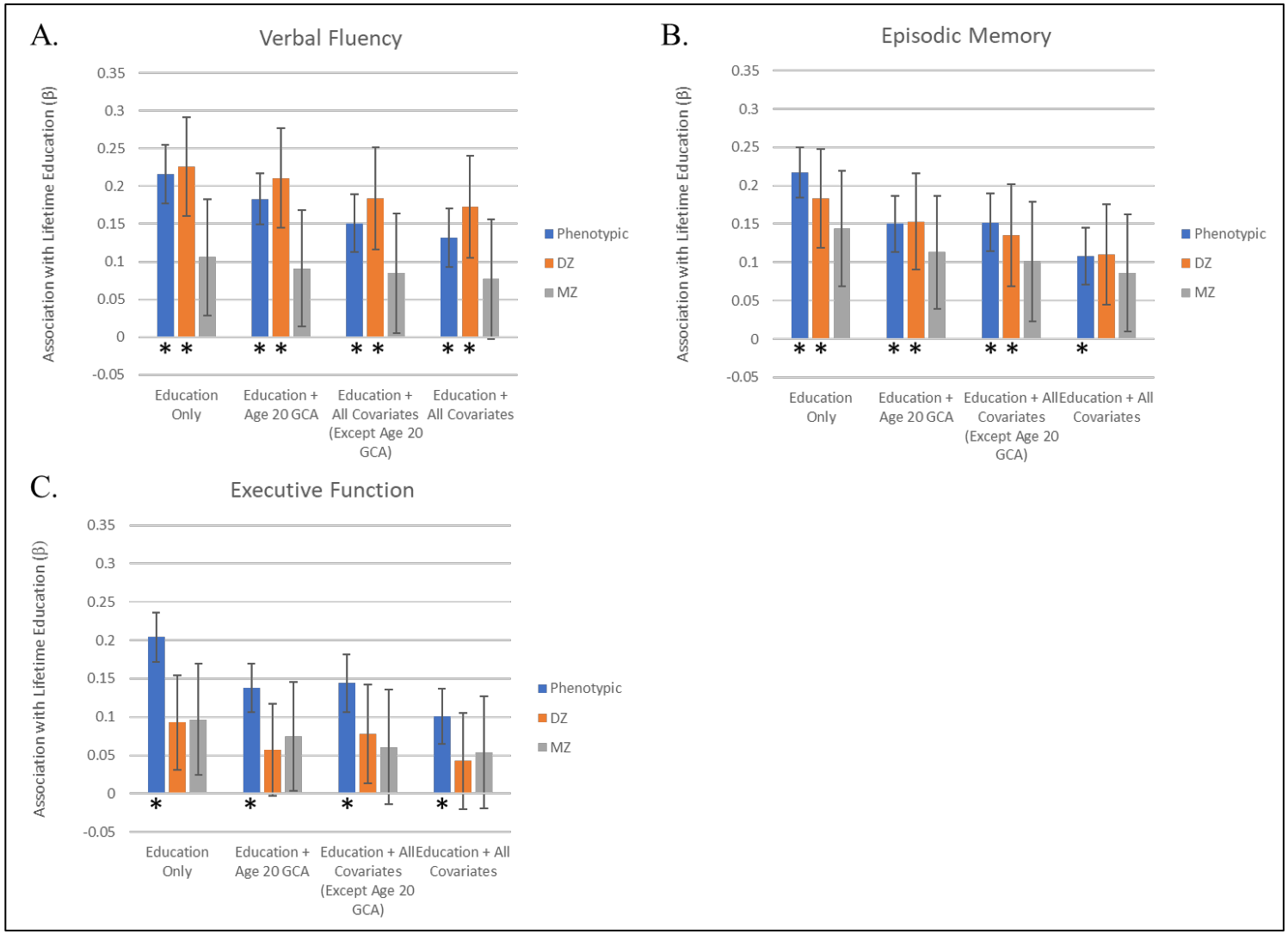
Supplementary Results

Sample

Characteristics. The median occupational complexity classification was 3341 (Interquartile Range, 2421-7222; Range, 0-9623). As noted in *Materials and Methods*, these numbers refer to specific occupational codes. The median of 3341 indicates technician and associate professionals (3), business and administration associate professionals (3), administrative and specialized secretaries (4), office supervisors (1). The first digit "2" of the upper end of the interquartile range indicates professionals, and the first digit "7" of the lower end indicates craft and related trade workers. Additional sample characteristics are shown in Tables S4 and S5.

Cotwin-Control. Using the cotwin-control approach, we examined the effect of lifetime education on selected cognitive abilities after accounting for age 20 GCA and other predictors (see Supplementary Figure S1). We used this approach for verbal fluency, episodic memory, and executive function because those showed larger effects of education, although still only about 1% of the variance in these tasks was attributable to lifetime education after accounting for the other predictors. The effect of lifetime education on verbal fluency was reduced by half among MZ pairs and was nonsignificant in the fully adjusted cotwin-control model compared to DZ pairs or the regression in the total sample. This result is supportive of partial genetic/familial confounding. A similar pattern was observed for executive functioning. However, the effect for episodic memory suggested that the MZ within-pair effect was comparable in effect size to the total sample and to the within-DZ pair effect. This latter pattern would be suggestive of a direct environmental effect. However, the effect is very small in nature and nonsignificant as it reflects only a part of about 1% of the variance. **Cotwin-Control.** Using the cotwin-control approach, we

examined the effect of lifetime education on selected cognitive abilities after accounting for age 20 GCA and other predictors (see Supplementary Figure S1). We used this approach for verbal fluency, episodic memory, and executive function because those showed larger effects of education, although still only about 1% of the variance in these tasks was attributable to lifetime education after accounting for the other predictors. The effect of lifetime education on verbal fluency was reduced by half among MZ pairs and was nonsignificant in the fully adjusted cotwin-control model compared to DZ pairs or the regression in the total sample. This result is supportive of partial genetic/familial confounding. A similar pattern was observed for executive functioning. However, the effect for episodic memory suggested that the MZ within-pair effect was comparable in effect size to the total sample and to the within-DZ pair effect. This latter pattern would be suggestive of a direct environmental effect. However, the effect is very small in nature and nonsignificant as it reflects only a part of about 1% of the variance. Finally, because the effect of education could be smaller in the present study than in other studies because of our inclusion of several additional covariates in the models, we also examined models with only age 20 GCA and lifetime education as predictors. Results of these models are depicted in the set of bars that are second to the left in each panel of Figure S1. As can be seen in the figure, the pattern of results of these cotwin-control analyses was very similar to that of the other analyses.



Supplementary Figure S1. Co-twin control analyses depicting the effect of education on verbal fluency (A), episodic memory (B), or executive function (C). We plotted the regression effect of education at the individual level (phenotypic), as well as the within dizygotic (DZ) and monozygotic (MZ) twin pairs discordant for education (measured continuously). Analyses were conducted with education alone in the model (left), education and age 20 GCA (second from left), all covariates except age 20 GCA (second from right), and all covariates including age 20 GCA (right). Error bars denote 1 standard error.
* Indicates the effect of education was statistically significant ($p<.05$).

Table S1. Correlations Among Predictor Variables

	Childhood SES	Age 20 GCA	Lifetime Education	Occupational Complexity	Engagement in Cognitive Activities	Physical Activity
Age 20 GCA	0.178 p<.0001					
Lifetime Education	0.298 p<.0001	0.274 p<.0001				
Occupational Complexity	0.157 p<.0001	0.133 p<.0001	0.410 p<.0001			
Engagement in Cognitive Activities	0.195 p<.0001	0.206 p<.0001	0.417 p<.0001	0.254 p<.0001		
Physical Activity	0.059 p=0.066	0.018 p=0.5628	0.246 p<.0001	0.124 p<.0001	0.342 p<.0001	
Health Status	-0.045 p=0.1653	-0.030 p=0.3482	-0.040 p=0.2009	0.043 p=0.1761	-0.067 p=0.0322	-0.084 p=0.0077

Note: SES, Socioeconomic status; GCA, General cognitive ability. Engagement in cognitive activities, physical activity, and health status were assessed at average age 56.

Table S2. Neuropsychological Test Battery

Cognitive Domain	Test Instrument	Scores
GCA	AFQT	Percentile (transformed to normal deviates for analyses)
Abstract Reasoning	WASI Matrix Reasoning	
Episodic Memory	CVLT-II	Sum of trials 1-5, short-delay free recall, long-delay free recall
		Logical Memory subtest immediate recall, delayed recall
		Visual Reproductions subtest immediate recall, delayed recall
	WMS-III	
Processing Speed	D-KEFS	Trail Making Test number sequencing, letter sequencing (times)
		Word reading, color naming (number of correct words for each in 45 seconds)
	Stroop Test	
Verbal Fluency	D-KEFS	Phonemic fluency (total words generated for F, A, and S)
		Category fluency (total words generated for animals and boys' names—1 minute each)
Visual-Spatial Ability	Gottschaldt Hidden Figures	Total number correct
	Card Rotation	Total number correct
	WMS-III	Visual Reproductions copy
Short-Term/Working Memory	WMS-III	Digit Span, Spatial Span, Letter-Number Sequencing
		Trail Making Test cancellations (number correct)
	D-KEFS	
Executive Function	Inhibition	Color-Word adjusted for word reading and color naming conditions
	Stroop	(number of correct words in 45 seconds)
		d' prime
	AX-CPT	Trail Making Test (switching condition time adjusted for number sequencing and letter sequencing conditions)
	Shifting	Category Switching (Fruits and Furniture switching accuracy
	D-KEFS	adjusted for animals and boys' names conditions)
	Working Memory Span	Letter-Number Sequencing
	WMS-III	Digit Span (total number of correct trials across forward and backward conditions)
	Reading Span	Total number of correct words recalled

Table S3. Correlations Among Outcome Variables

	Age 62 GCA	Abstract Reasoning	Episodic Memory	Processing Speed	Verbal Fluency	Visual-Spatial Ability	Short-Term/Working Memory
Abstract Reasoning	0.565 p<.0001						
Episodic Memory	0.467 p<.0001	0.384 p<.0001					
Processing Speed	0.426 p<.0001	0.425 p<.0001	0.352 p<.0001				
Verbal Fluency	0.294 p<.0001	0.281 p<.0001	0.372 p<.0001	0.471 p<.0001			
Visual-Spatial Ability	0.617 p<.0001	0.511 p<.0001	0.359 p<.0001	0.479 p<.0001	0.301 p<.0001		
Short-Term/Working Memory	0.507 p<.0001	0.438 p<.0001	0.424 p<.0001	0.526 p<.0001	0.423 p<.0001	0.451 p<.0001	
Executive Function	0.523 p<.0001	0.459 p<.0001	0.478 p<.0001	0.506 p<.0001	0.417 p<.0001	0.4709 p<.0001	0.752 p<.0001

GCA, General cognitive ability.

Table S4. Sample Characteristics, Part 1

Characteristics	Mean (Standard Deviation)	Range
Age at Young Adult GCA Assessment (y)	19.79 (1.29)	17.05-25.66
Age at VETSA 1 (y)	55.90 (2.44)	51.08-60.67
Age at VETSA 2: Outcome Assessment: (y)	61.72 (2.45)	56.50-66.50
Young Adult GCA (percentile)*	61.48 (22.07)	10-99
Engagement in Cognitive Activities†	3.36 (2.39)	0-11
Physical Activity†	2.59 (1.20)	1-5
Health Status‡	1.06 (1.15)	0-6
Education at Young Adult GCA Assessment (y)	12.30 (1.21)	7-20
Lifetime Education (y)	13.88 (2.09)	8-20

GCA, General cognitive ability

* Based on AFQT (percentiles were transformed in the analyses in order to normalize the distribution).

† Number of activities.

‡ Number of chronic medical conditions: diabetes, emphysema, asthma, cancer, osteoarthritis, rheumatoid arthritis, stroke, heart attack, heart failure, heart surgery, angina, hypertension, peripheral vascular disease, cirrhosis, AIDS.

Table S5. Sample Characteristics, Part 2

Characteristics	Number of Subjects (%)
Lifetime Education < 12 y	28 (2.78%)
Lifetime Education = 12 y	380 (37.66%)
Lifetime Education = 13-15 y	307 (30.43%)
Lifetime Education ≥ 16 y	294 (29.13%)
Race/ethnicity (white, non-Hispanic)	903 (89.49%)

Table S6. Model 2: Predictors of Late Midlife (Average Age 62) Cognitive Function Including Age 20 Education and Lifetime Education

Cognitive Ability/Domain	Childhood SES	Age 20 Education	Lifetime Education	Occupational Complexity	Engagement in Cognitive Activities	Physical Activity	Health Status
Age 62 GCA (n=955)	.039 p=.304	.051 p=.155	**.108 p=.005**	.034 p=.293	**.124 p=.00029**	-.057 p=.072	**-.073 p=.017**
Abstract Reasoning (n=956)	.085 p=.018	.061 p=.082	**.099 p=.011**	**.079 p=.016**	**.106 p=.002**	**-.080 p=.014**	**-.099 p=.001**
Episodic Memory (n=957)	.009, p=.809	.102 p=.005	**.114 p=.003**	**.078 p=.016**	**.086 p=.012**	-.030 p=.345	-.036 p=.238
Processing Speed (n=954)	.038, p=.303	.046 p=.206	.069 p=.084	**.118 p=.00042**	**.100 p=.005**	.019 p=.563	**-.081 p=.010**
Verbal Fluency (n=955)	-.001 p=.979	.025 p=.495	**.150 p=.00011**	.068 p=.036	**.138 p=.000055**	-.001 p=.982	.002 p=.953
Visual-Spatial Ability (n=948)	.063, p=.095	.048 p=.178	**.094 p=.014**	.061 p=.057	**.127 p=.0002**	-.030 p=.342	**-.076 p=.013**
Short-Term/Working Memory (n=956)	.045 p=.239	.078 p=.032	**.123 p=.001**	**.086 p=.007**	.031 p=.358	-.015 p=.631	-.033 p=.279
Executive Function (n=958)	.050 p=.180	.080 p=.025	**.122 p=.001**	**.101 p=.001**	.029 p=.378	-.030 p=.328	-.063 p=.033

Note: SES, Socioeconomic status; GCA, General cognitive ability. Engagement in cognitive activities, physical activity, and health status were assessed at average age 56. All outcomes were adjusted for age and race/ethnicity. Numbers in the table are β coefficients. Numbers in **bold** are significant after correction for multiple testing. Exact p-values are shown to further highlight differences in magnitude of effects.

Table S7. Model 5: Predictors of Late Midlife (Average Age 62) Cognitive Function Including Age 20 GCA and Lifetime Education (12 Years vs. 16+ Years)—Based on Comparison with Clouston et al. (1)

	Childhood SES	Age 20 GCA	Lifetime Attained Education (Dichotomized)
Age 62 GCA *n*=463	-.037 p=.346	**.617** **p=8.3e-52**	**.106** **p=.007**

Note: SES, Socioeconomic status; GCA=General cognitive ability. Engagement in cognitive activities, physical activity, and health status were assessed at average age 56. All outcomes were adjusted for age and race/ethnicity. Numbers in the table are β coefficients. Numbers in **bold** are significant. There were 333 individuals with 12 years of lifetime education and 154 with a lifetime university (4-year college) education. Sample size for this analysis includes 95% of the total n of 487 (i.e., 333+154). Exact p-values are shown to highlight differences in magnitude of effects.

Table S8. Models Including Only a Single Factor in Addition to Age 20 GCA

	Model 6a			Model 6b	
Cognitive Ability/Domain	Age 20 GCA	Lifetime Education		Age 20 GCA	Occupational Complexity
Age 62 GCA	**.644** **p=3.35e-88**	.046 p=.067		**.650** **p= 2.12e-92**	.044 p=.064
Abstract Reasoning	**.376** **p=5.41e-32**	**.121** **p=8.37e-05**		**.391** **p=9.80e-36**	**.113** **p=.00011**
Episodic Memory	**.295** **p=7.98e-20**	**.154** **p=9.16e-07**		**.318** **p=2.68e-23**	**.120** **p=3.93e-05**
Processing Speed	**.217** **p=7.14e-11**	**.138** **p=2.67e-05**		**.232** **p=1.11e-12**	**.157** **p=3.67e-07**
Verbal Fluency	**.148** **p=6.93e-06**	**.198** **p=1.12e-09**		**.179** **p=4.10e-08**	**.138** **p=6.03e-06**
Visual-Spatial Ability	**.424** **p=1.04e-41**	**.111** **p=.00019**		**.437** **p=1.13e-45**	**.100** **p=.00028**
Short-Term/Working Memory	**.333** **p=7.97e-26**	**.136** **p=9.13e-06**		**.350** **p=3.25e-29**	**.119** **p=2.44e-05**
Executive Function	**.340** **p=4.12e-27**	**.146** **p=1.50e-06**		**.358** **p=6.86e-31**	**.136** **p=1.51e-06**

Note: GCA=General cognitive ability. All outcomes were adjusted for age and race/ethnicity. Numbers in the table are β coefficients. Numbers in **bold** are significant after correction for multiple testing. Exact p-values are shown to highlight differences in magnitude of effects.

1. Clouston SA*, et al.* (2012) Benefits of educational attainment on adult fluid cognition: International evidence from three birth cohorts. *Int. J. Epidemiol.* 41:1729-1736.

References

1. International Labour Office (2012) *International Standard Classification of Occupations: Structure, Group Definitions and Correspondence Tables*. ISCO-08 (International Labour Office, Geneva, Switzerland), Vol 1.
2. Beck A*, et al.* (2018) Mediators of the effect of childhood socioeconomic status on late midlife cognitive abilities: A Four decade longitudinal study. *Innov Aging* 2(1):igy003-igy003.
3. Franz CE*, et al.* (2011) A 35-year longitudinal assessment of cognition and midlife depression symptoms: The Vietnam Era Twin Study of Aging. *Am. J. Geriatr. Psychiatry* 19:559-570.
4. Lezak MD, Howieson DB, Bigler ED, & Tranel D (2012) *Neuopsychological assessment* (Oxford University Press, New York) 5th Ed.
5. Gustavson DE*, et al.* (2018) Genetic and environmental architecture of executive functions in midlife. *Neuropsychology* 32(1):18-30.
6. Lyons MJ*, et al.* (2009) Genes determine stability and the environment determines change in cognitive ability during 35 years of adulthood. *Psychol. Sci.* 20:1146-1152.
7. Lyons MJ*, et al.* (2017) A longitudinal twin study of general cognitive ability over four decades. *Dev. Psychol.* 53(6):1170-1177.
8. Bayroff AG & Anderson AA (1963) Development of Armed Forces Qualification Tests 7 and 8 (Technical Research Report 1122). (U.S. Army Research Institute, Alexandria, VA).
9. Wechsler D (1997) *Manual for the Wechsler Adult Intelligence Scale - Third Edition* (Psychological Corporation, San Antonio, TX).
10. Delis DC, Kramer JH, Kaplan E, & Ober BA (2000) *California Verbal Learning Test* (Psychological Corporation, San Antonio, TX) 2nd Ed.
11. Wechsler D (1997) *Manual for the Wechsler Memory Scale - Third Edition* (Psychological Corporation, San Antonio, TX).
12. Golden CJ & Shawna MF (1978) Stroop color and word test.
13. Delis DC, Kaplan E, & Kramer JH (2001) *Delis-Kaplan Executive Function System (D-KEFS)* (Psychological Corporation, San Antonio, TX).
14. Ekstrom RB, French JW, & Harmon HH (1976) *Kit of Factor-Referenced Tests (Manual)* (Educational Testing Service, Princeton, NJ).
15. Thurstone LL (1944) *A Factorial Study of Perception* (University of Chicago Press, Chicago).
16. Daneman M & Carpenter PA (1980) Individual-Differences in Working Memory and Reading. *J Verb Learn Verb Be* 19(4):450-466.
17. Friedman NP*, et al.* (2008) Individual differences in executive functions are almost entirely genetic in origin. *J. Exp. Psychol. Gen.* 137:201-225.
18. Engelhardt LE, Briley DA, Mann FD, Harden KP, & Tucker-Drob EM (2015) Genes unite executive functions in childhood. *Psychol. Sci.* 26(8):1151-1163.
19. Kremen WS*, et al.* (2010) Genetic and environmental influences on the size of specific brain regions in midlife: The VETSA MRI study. *Neuroimage* 49:1213-1223.
20. Eyler LT*, et al.* (2011) Genetic and environmental contributions to regional cortical surface area in humans: A magnetic resonance imaging twin study. *Cereb. Cortex* 21(10):2313-2321.
21. Elman JA*, et al.* (2017) Genetic and environmental influences on cortical mean diffusivity. *Neuroimage* 146:90-99.
22. Hatton SN*, et al.* (2018) Negative fateful life events in midlife and advanced predicted brain aging. *Neurobiol. Aging* 67:1-9.

23. Dale AM, Fischl B, & Sereno MI (1999) Cortical surface-based analysis. I: Segmentation and surface reconstruction. *Neuroimage* 9:179-194.
24. Fischl B & Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U.S.A.* 97:11050-11055.
25. Fischl B, Sereno MI, & Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9:195-207.
26. Jovicich J*, et al.* (2006) Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *Neuroimage* 30:436-443.
27. Benjamini Y & Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 57:289-300.
28. Li J & Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95(3):221-227.
29. Clouston SA*, et al.* (2012) Benefits of educational attainment on adult fluid cognition: International evidence from three birth cohorts. *Int. J. Epidemiol.* 41:1729-1736.
30. Carlin JB, Gurrin LC, Sterne JA, Morley R, & Dwyer T (2005) Regression models for twin studies: a critical review. *Int J Epidemiol* 34(5):1089-1099.
31. McGue M, Osler M, & Christensen K (2010) Causal inference and observational aging research: The utility of twins. *Perspectives on Psychological Science* 5(5):546-556.