ELSEVIER

Neurotoxicology and Teratology 23 (2001) 569–589

NEUROTOXICOLOGY AND TERATOLOGY

Neurobehavioral test performance in the third National Health and Nutrition Examination Survey☆

Edward F. Krieg, Jr.[a,*], David W. Chrislip[a], Richard E. Letz[b], David A. Otto[c], Carlos J. Crespo[d], W. Stephen Brightwell[a], Richard L. Ehrenberg[e]

[a]*National Institute for Occupational Safety and Health, Robert A. Taft Laboratories, 4676 Columbia Parkway, MS C-22, Cincinnati, OH 45226, USA*
[b]*Emory University, Atlanta, GA, USA*
[c]*U.S. Environmental Protection Agency, Research Triangle Park, NC, USA*
[d]*State University of New York, Buffalo, NY, USA*
[e]*National Institute for Occupational Safety and Health, Atlanta, GA, USA*

Received 3 April 2000; accepted 6 September 2001

Abstract

The third National Health and Nutrition Examination Survey (NHANES III) contained three computerized neurobehavioral tests from the Neurobehavioral Evaluation System (NES): simple reaction time, symbol–digit substitution and serial digit learning. The neurobehavioral data that were collected came from a nationally representative sample of adults 20–59 years old. Performance on the tests was related to sex, age, education level, family income and race-ethnicity. Performance decreased as age increased, and increased as education level and family income increased. Differences in performance between sexes, levels of education and racial-ethnic groups tended to decrease as family income increased. The relationship between age and performance on the symbol–digit substitution test varied by education level and by racial-ethnic group. The relationship between age and performance on the serial digit learning test varied by racial-ethnic group. Questionnaire variables that were related to performance on one or more of the tests included the reported amount of last night's sleep, energy level, computer or video game familiarity, alcoholic beverages within the last 3 h and effort. Persons who took the tests in English or Spanish performed differently on the symbol–digit substitution and serial digit learning tests. Performance on all the tests decreased as test room temperature increased. Published by Elsevier Science Inc. All rights reserved.

Keywords: NES; NHANES III; Simple reaction time; Symbol–digit substitution; Serial digit learning; Demographic; Questionnaire

1. Introduction

The National Center for Health Statistics periodically conducts large surveys to assess the health and nutrition of persons in the United States. The surveys consist of questionnaires, physical examinations and laboratory tests. The components of the surveys are administered in homes and in mobile examination centers. The surveys have provided national reference data for biological markers and anthropometric measurements (e.g., Refs. [20,21,32,55,61]). The last completed survey was the third National Health and Nutrition Examination Survey (NHANES III). This survey's documentation [49] and data [50,51] are available on CD-ROM.

NHANES III contained three neurobehavioral tests from a computerized test battery called the Neurobehavioral Evaluation System (NES) that were used to assess central nervous system (CNS) function. The three tests were simple reaction time, symbol–digit substitution and serial digit learning. They are often referred to as the CNS component of the survey. The NES was designed to measure neurobehavioral performance in populations that are occupationally or environmentally exposed to chemicals [7,8,34–36,38,39]. It has been used to investigate the neurotoxic effects of organic solvents [6,13,15,19,24,27,48], organophosphate

☆ This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the agency.

* Corresponding author. Tel.: +1-513-533-8160; fax: +1-513-533-8510.

E-mail address: erk3@cdc.gov (E.F. Krieg Jr.).

PII: S0892-0362(01)00177-5

pesticides [47,63], nitrous oxide [26,46], styrene [40,64], volatile organic compounds [53], lead [60,65,69], mercury [22,23,44,65], tetrachloroethane [2] and sulfuryl fluoride and methyl bromide [16].

The purposes of this article are to describe the survey methods with respect to the neurobehavioral tests and to summarize the neurobehavioral data statistically using the demographic variables that were included in the survey and questionnaire items that were administered with the tests. There are no data for the NES from a large sample of the general population. The data collected by NHANES III are from a nationally representative sample. Estimates calculated from this sample can be used as reference values for workplace and other studies designed to explore the etiology of nervous system disorders. They provide a reference point against which future national surveys can be compared in order to evaluate trends over time. The estimates can be used for power and sample size calculations, and they can be used as a guide for deciding which variables to include and control for in future studies.

2. Methods

2.1. Subjects

The subjects in NHANES III were civilian, non-institutionalized persons in the United States 2 months of age or older. The survey was conducted from 1988 to 1994. Approximately 40,000 persons were selected to participate.

Table 1
Test parameters

Value	Description
Simple reaction time	
5	Number of blocks with preferred hand
0	Number of blocks with nonpreferred hand
10	Number of trials per block
2500	Minimum interstimulus interval in milliseconds
2500	Maximum for variable part of ISI in milliseconds
1000	Stimulus interval and maximum RT in milliseconds
6	Number of 'errors' per block before error message
7	Index for color of stimulus
2	ITI mode
Symbol–digit substitution	
4	Number of trials
1	Number of practice trials
1	Error entry mode
4	Number of errors allowed before 'call the interviewer'
1	Stimulus mode
1	Symbol–digit pairing over trials
0	Number of incidental learning trials
176,3,238,239,247,251,254,215,127, * nine character codes	
Serial digit learning	
600	Stimulus interval in milliseconds
600	Interstimulus interval in milliseconds
1	Stimulus mode
8	Maximum number of trials to be administered
8	Span length
4	Span length of practice trial

Table 2
Proportion of persons who did not complete the tests by demographic categories

Demographic variable	*n*	Simple reaction time	Symbol–digit substitution	Serial digit learning
Overall	5662	0.0925	0.1007	0.1236
Sex				
Male	2594	0.0968	0.1072	0.1295
Female	3068	0.0890	0.0952	0.1186
Age				
20–29	1784	0.0891	0.0908	0.1076
30–39	1616	0.0910	0.0965	0.1132
40–49	1327	0.0934	0.0980	0.1372
50–59	935	0.1005	0.1305	0.1529
Last grade attended				
0–8	931	0.1568	0.1987	0.2739
9–11	973	0.0915	0.0966	0.1357
12	1871	0.0764	0.0775	0.0844
13+	1851	0.0773	0.0773	0.0810
Family income				
<10,000	856	0.1133	0.1343	0.1741
10,000–29,999	2291	0.0904	0.0969	0.1248
30,000–49,999	1223	0.0809	0.0834	0.0924
≥50,000	864	0.0648	0.0660	0.0775
Race-ethnicity				
Non-Hispanic white	1924	0.0551	0.0582	0.0639
Non-Hispanic black	1791	0.1033	0.1122	0.1351
Mexican American	1708	0.1048	0.1189	0.1610
Other	239	0.2259	0.2259	0.2510

Age is in years.
Family income is in dollars.

Table 3
Percentiles of the test measures

	Simple reaction time	Symbol–digit substitution		Serial digit learning	
Percentile	Mean reaction time	Mean total latency	Number of errors	Trials to criterion	Total score
1	175.86	14.62			
5	186.42	16.15			
10	193.06	16.98			
20	202.35	18.11		2.22	0.36
30	209.62	19.04		2.63	1.04
40	216.61	20.06		3.05	1.68
50	224.08	21.00		3.62	2.47
60	231.48	22.17	0.31	4.28	3.47
70	240.75	23.70	0.76	5.26	4.91
80	253.27	25.95	1.44	6.88	7.39
90	276.84	30.14	2.85		11.54
95	312.43	34.94	4.83		14.19
99	450.67	49.64	9.54		

Mean reaction time is in milliseconds.
Mean total latency is in seconds.

Table 4
Performance on the simple reaction time test by demographic categories

Demographic variable	*n*	Mean reaction time		
		Mean	S.E.	DEFF
Overall	4896	233.46	1.38	4.21
Sex				
Male	2234	226.18	1.78	3.57
Female	2662	240.49	1.92	4.13
Age				
20–29	1526	230.70	1.91	2.78
30–39	1407	231.37	1.83	2.32
40–49	1151	234.40	2.65	3.73
50–59	812	240.83	2.41	1.58
Last grade attended				
0–8	720	257.64	5.12	3.58
9–11	834	243.73	3.72	3.93
12	1656	235.70	2.02	2.96
13+	1659	225.23	1.57	3.10
Family income				
<10,000	699	247.52	3.64	1.98
10,000–29,999	1961	237.55	1.89	2.71
30,000–49,999	1097	232.06	2.46	3.49
≥50,000	798	224.31	1.62	1.97
Race-ethnicity				
Non-Hispanic white	1795	229.90	1.50	2.42
Non-Hispanic black	1552	247.59	2.56	2.21
Mexican American	1368	247.69	1.97	1.59
Other	181	239.48	4.63	1.22

Mean reaction time is in milliseconds.
Age is in years.
Family income is in dollars.
S.E.: standard error.
DEFF: design effect.

Of these, 5662 adults, 20–59 years old, were selected to take the neurobehavioral tests. Persons were not excluded from the CNS test for medical or safety reasons. Persons who could not speak English or Spanish, or who were legally blind were not given the CNS test.

2.2. Sample design

The sample design was a stratified, multistage probability design. In the first stage of sampling, 81 primary sampling units (PSUs) were selected. The PSUs were individual counties or adjacent counties. Thirteen of the large PSUs were divided into 21 survey locations and the remaining 68 PSUs had one survey location. The 89 survey locations or "stands" were randomly divided into two phases. Phase I consisted of 44 locations visited from 1988 to 1991. Phase II consisted of 45 locations visited from 1991 to 1994.

In the second stage of sampling, 2144 area segments were selected. The area segments were usually city or town blocks, or groups of blocks. In the third stage of sampling, 19,528 households and certain types of group living quarters were selected. In the fourth stage, 39,695 sample persons were selected to be interviewed in their homes and to be given exams in a mobile examination center. The number of persons actually interviewed at home was 33,994, and 30,818 participated in an examination at a mobile examination center. Young children, older persons, black persons

Table 5
Performance on the simple reaction time test by questionnaire categories

Questionnaire variable	*n*	Mean reaction time		
		Mean	S.E.	DEFF
Last night's sleep				
Less than usual	1467	235.19	1.88	2.26
About usual amount	2929	233.67	1.65	3.54
More than usual	437	223.91	2.31	1.74
Energy Level				
Energetic	259	225.69	2.38	0.90
Fresh	626	230.94	2.84	2.15
Average	2660	230.81	1.51	3.28
Tired	1146	238.28	1.98	1.92
Exhausted	143	263.78	11.87	2.72
Computer or video game familiarity				
None	2002	244.75	2.39	3.47
Some	2148	229.00	1.49	2.69
A lot	684	226.30	2.30	2.96
Drinks with caffeine in last 3 h				
0	4635	233.11	1.36	3.97
1	136	233.11	4.86	1.63
2+	63	239.02	8.22	0.81
Drinks with alcohol in last 3 h				
0	4794	233.12	1.34	3.93
1	24	252.53	19.64	1.39
2+	16	241.66	12.43	0.68
Preferred hand				
Right	4434	233.40	1.36	3.73
Left	400	230.80	2.94	1.63
Effort				
None	119	240.76	4.83	0.97
Some	1189	232.25	1.56	1.63
A lot	1081	233.17	2.79	3.48
As hard as I could	2443	233.32	1.41	2.18
Test language				
English	4175	232.01	1.30	3.54
Spanish	659	260.62	5.58	3.59
Test room temperature				
15.6–19.4	30	227.96	8.08	1.79
20.0–21.1	286	227.89	2.87	1.67
21.7–22.8	985	233.86	1.97	2.49
23.3–24.4	1609	233.66	2.24	2.97
25.0–26.1	887	237.78	2.84	2.68
26.7–27.8	170	254.75	7.09	2.36
28.3–33.3	10	263.11	14.09	0.59

Mean reaction time is in milliseconds.
Test room temperature is in degrees Celsius.
S.E.: standard error.
DEFF: design effect.

Table 6
Summary table for mean reaction time

Variable	*b*	S.E.	DEFF	*df*	F	P	SSPCC
Sex				1	23.89	.0000	0.017251
Age	0.21	0.09	1.29	1	5.45	.0237	0.001718
Last grade attended				3	2.93	.0428	0.008839
Family income				3	3.05	.0370	0.004060
Race-ethnicity				3	5.48	.0025	0.005644
Family income × Race-ethnicity				9	2.88	.0082	0.005198
Last night's sleep				2	4.27	.0195	0.003232
Energy level				4	2.75	.0385	0.007501
Computer or video game familiarity				2	2.66	.0801	0.003359
Drinks with caffeine in last 3 h				2	1.41	.2549	0.000908
Drinks with alcohol in last 3 h				2	0.65	.5265	0.000172
Preferred hand				1	0.04	.8509	0.000009
Effort				3	0.97	.4149	0.001494
Test language				1	2.76	.1028	0.002377
Test room temperature	1.48	0.71	2.17	1	4.37	.0418	0.002321
Phase	3.34	1.80	1.30	1	3.44	.0699	0.001094

Denominator $df = 49$.
$R^2 = .105118$.
b: slope.
S.E.: standard error.
DEFF: design effect.
SSPCC: squared semipartial correlation coefficient.

and Mexican Americans were over-sampled in order to produce more precise estimates for these groups.

The subjects selected for CNS testing were a half-sample of persons 20–59 years old who had an exam at a mobile examination center (MEC). The persons were systematically selected, those with odd-numbered survey identification numbers participated in the CNS test and those with even numbers participated in allergy testing. Subjects selected for the CNS test were randomly assigned to different technicians in order to reduce any systematic bias attributable to the manner in which individual technicians administered the test. There were 40 technicians who tested from 1 to 435 participants each.

2.3. Technician training

Training in the administration of the CNS test was given to an initial cohort of technicians during a 2-day period prior to the initiation of data collection. The training started with classroom instruction, which included the purpose of the test, background on the test's development, the importance of strict adherence to the standardized protocol, a description of each neurobehavioral test, step-by-step instructions for the administration of each neurobehavioral test and questionnaire item, criteria for aborting the test and procedures for recording and transmitting data. The training also included practice administering the test, first on other MEC staff and then on paid volunteers.

Similar training sessions were conducted for each new technician as she or he was hired or transferred into a position that entailed administering the test. Annual retraining sessions were conducted for each technician who administered the test.

Table 7
Performance on the simple reaction time test by family income and race-ethnicity

		Mean reaction time		
Family income race-ethnicity	*n*	Mean	S.E.	DEFF
< 10,000				
Non-Hispanic white	132	236.89	5.36	1.02
Non-Hispanic black	313	269.62	5.69	1.37
Mexican American	223	254.81	3.84	0.69
Other	31	247.29	10.64	1.99
10,000–29,999				
Non-Hispanic white	550	233.37	2.29	1.63
Non-Hispanic black	685	246.63	2.89	1.32
Mexican American	651	249.27	2.36	1.10
Other	75	244.82	9.16	1.25
30,000–49,999				
Non-Hispanic white	518	231.15	2.91	2.51
Non-Hispanic black	306	237.32	3.16	0.91
Mexican American	232	239.18	2.89	0.86
Other	41	232.85	5.81	0.85
≥ 50,000				
Non-Hispanic white	511	223.80	1.73	1.56
Non-Hispanic black	133	227.80	4.61	1.34
Mexican American	131	236.07	6.36	3.21
Other	23	225.75	7.17	1.04

Mean reaction time is in milliseconds.
Family income is in dollars.
S.E.: standard error.
DEFF: design effect.

2.4. Description of the neurobehavioral tests

The three neurobehavioral tests that were administered were components of the Neurobehavioral Evaluation System 2 (NES2, Neurobehavioral Systems, Winchester, MA). The parameters used for each test are given in Table 1.

2.4.1. Simple reaction time

The simple reaction time test is a test of visuomotor speed. The subjects were instructed to rest the index finger of their preferred hand on a push button and press the button as quickly as possible whenever a solid square (4×4 cm) was displayed in the center of a blank computer screen. The square remained on the screen until the subject pressed the button, after which it was cleared from the screen. If there was no response, the square was cleared from the screen after an elapsed time of 1000 ms. A total of 50 trials were administered to each subject. The interval between trials varied randomly according to a uniform distribution ranging from 2.5 to 5.0 s in order to limit anticipatory responses. The measured response was the latency, in milliseconds, between the appearance of the square on the screen and the subject pressing the button.

2.4.2. Symbol–digit substitution

The symbol–digit substitution test measures coding speed. On the upper half of the computer screen, the subject was presented with a grid that paired one of nine different symbols with one of the digits from 1 to 9. A similar grid was displayed on the bottom half of the screen with the same symbols presented in a scrambled order and the spaces for the corresponding digits left blank. The subject's task was to enter, as quickly as possible, the matching digit for each symbol. A total of five trials were conducted, with a different pairing of digits and symbols on each trial. The first trial was a practice trial that did not accept incorrect responses. The remaining trials allowed incorrect responses, but more than four errors caused the display of a message stressing the need to avoid errors. The amount of time, in seconds, required to enter each digit and the number of errors on each trial were recorded.

Table 8
Performance on the symbol–digit substitution test by demographic categories

		Mean total latency			Number of errors		
Demographic variable	*n*	Mean	S.E.	DEFF	Mean	S.E.	DEFF
Overall	5092	22.78	0.21	4.51	1.18	0.05	3.30
Sex							
Male	2316	23.46	0.26	3.12	1.00	0.05	1.54
Female	2776	22.13	0.25	3.39	1.35	0.08	3.23
Age							
20–29	1622	20.48	0.26	3.70	0.88	0.07	2.55
30–39	1460	21.23	0.21	2.83	1.01	0.08	2.57
40–49	1197	24.35	0.31	1.85	1.39	0.09	1.94
50–59	813	27.48	0.42	1.76	1.72	0.12	1.63
Last grade attended							
0–8	746	33.49	0.73	2.53	2.25	0.18	2.55
9–11	879	26.16	0.42	2.15	1.37	0.11	1.77
12	1726	22.95	0.22	2.41	1.31	0.09	2.74
13+	1708	20.10	0.13	2.23	0.87	0.05	1.77
Family income							
<10,000	741	25.14	0.51	1.91	1.74	0.16	2.32
10,000–29,999	2069	23.67	0.29	2.99	1.23	0.09	3.40
30,000–49,999	1121	22.29	0.29	2.81	1.16	0.09	1.90
≥50,000	807	20.74	0.24	2.12	0.93	0.10	2.73
Race-ethnicity							
Non-Hispanic white	1812	21.71	0.22	3.18	1.04	0.07	2.16
Non-Hispanic black	1590	25.81	0.30	1.28	1.72	0.07	1.01
Mexican American	1505	27.66	0.46	2.38	1.77	0.09	1.64
Other	185	25.73	0.76	1.11	1.42	0.22	1.61

Mean total latency is in seconds.
Age is in years.
Family income is in dollars.
S.E.: standard error.
DEFF: design effect.

2.4.3. Serial digit learning

The serial digit learning test measures learning and recall. The subject was presented with a series of digits displayed one at a time on the computer screen. Each digit was displayed on the screen for 0.6 s with 0.6-s intervals between digits. After all the digits were displayed, the subject pressed the numeric keys on the keyboard to enter the entire sequence of numbers in the order in which they were presented. The first trial was a practice trial consisting of four digits. All the other trials contained the same eight-digit

Table 9
Performance on the symbol–digit substitution test by questionnaire categories

		Mean total latency			Number of errors		
Questionnaire variable	*n*	Mean	S.E.	DEFF	Mean	S.E.	DEFF
Last night's sleep							
Less than usual	1516	22.97	0.31	2.44	1.15	0.10	3.63
About usual amount	3058	22.76	0.23	3.41	1.20	0.07	2.89
More than usual	456	22.13	0.47	2.50	1.19	0.09	1.03
Energy level							
Energetic	275	22.31	0.40	1.43	1.16	0.15	1.71
Fresh	649	22.42	0.39	2.43	1.00	0.11	2.00
Average	2767	22.67	0.22	2.67	1.11	0.07	3.02
Tired	1196	22.95	0.33	2.41	1.42	0.13	3.87
Exhausted	144	25.98	1.36	2.00	1.52	0.31	1.96
Computer or video game familiarity							
None	2091	27.51	0.32	2.54	1.73	0.10	3.02
Some	2241	21.36	0.15	2.40	1.04	0.07	3.04
A lot	699	18.93	0.18	1.18	0.71	0.08	2.15
Drinks with caffeine in last 3 h							
0	4806	22.71	0.21	4.44	1.18	0.06	3.32
1	160	24.67	1.07	1.45	1.31	0.26	2.33
2+	65	22.99	0.74	1.28	1.18	0.23	1.67
Drinks with alcohol in last 3 h							
0	4972	22.73	0.21	4.53	1.18	0.05	3.25
1	44	33.31	2.61	0.75	3.49	0.72	1.92
2+	15	26.95	1.33	0.52	1.17	0.18	0.34
Preferred hand							
Right	4618	22.74	0.22	4.47	1.17	0.06	3.18
Left	413	22.96	0.36	1.21	1.31	0.15	1.98
Effort							
None	119	23.34	1.06	1.57	1.20	0.23	1.42
Some	1248	22.99	0.34	2.90	1.24	0.11	2.97
A lot	1118	21.97	0.28	2.51	1.10	0.07	1.38
As hard as I could	2545	23.04	0.22	2.20	1.20	0.07	2.36
Test language							
English	4368	22.30	0.21	4.54	1.15	0.06	3.08
Spanish	663	33.58	0.77	2.59	1.95	0.19	3.95
Test room temperature							
15.6–19.4	30	20.52	1.33	1.70	0.79	0.30	1.35
20.0–21.1	282	21.71	0.41	1.27	1.17	0.11	1.09
21.7–22.8	986	22.29	0.37	2.95	1.10	0.11	2.65
23.3–24.4	1598	22.37	0.27	2.54	1.17	0.08	2.53
25.0–26.1	882	23.89	0.40	2.12	1.13	0.09	1.89
26.7–27.8	170	25.71	0.88	1.77	1.51	0.14	0.54
28.3–33.3	10	27.62	2.22	0.45	1.65	0.25	0.54

Mean total latency is in seconds.
Test room temperature is in degrees Celsius.
S.E.: standard error.
DEFF: design effect.

sequence. Testing continued until the subject responded correctly on two consecutive trials or until the subject attempted eight trials. The sequence of digits entered by the subject was recorded for each trial.

2.5. Computer equipment

The tests were administered on a Compaq 286 DeskPro personal computer (12.5 MHz 80286 CPU chip) configured with 640K RAM, one 1.2 MB floppy disk drive, one parallel port, two serial ports, a standard game joystick controller board and an Ethernet interface board. The system used the standard detachable keyboard with a custom-made overlay that left only the top row of digit keys exposed, and a standard Compaq monochrome (green) monitor. In September 1991, the monitors were replaced with comparable, new (amber) monitors because of deterioration in display performance. This change occurred between the conclusion of Phase I data collection and the initiation of Phase II data collection. A standard computer game joystick (Model 100FJ, CH Products, Vista, CA) with push buttons was used to input subjects' simple reaction time test responses.

The Compaq personal computers were networked to a VAX system. The CNS testing programs were stored on the VAX and downloaded to the Compaq when the tests were administered. The VAX would initiate running the testing programs on the Compaq and store the response data. When the VAX or the connection to the VAX went down, testing programs stored on a floppy disk were used and the response data were stored on the floppy disk and then uploaded to the VAX system.

2.6. Test room

The neurobehavioral tests were administered in the audiometry room of a MEC. This room was soundproof and was located in an isolated part of the fourth trailer in order to minimize distractions from traffic in the hallways. The door to the examination room was closed when a

Table 10
Summary table for mean total latency

Variable	*b*	S.E.	DEFF	*df*	F	P	SSPCC
Sex				1	41.72	.0000	0.011283
Age	0.23	0.01	1.56	1	427.65	.0000	0.098872
Last grade attended				3	90.37	.0000	0.057375
Family income				3	12.51	.0000	0.004468
Race-ethnicity				3	47.93	.0000	0.020167
Sex × Last grade attended				3	1.05	.3794	0.000820
Sex × Race-ethnicity				3	4.84	.0050	0.001623
Age × Last grade attended				3	20.51	.0000	0.007115
0–8	0.40	0.03	0.98	1	158.76	.0000	
9–11	0.32	0.04	2.81	1	75.34	.0000	
12	0.22	0.02	1.40	1	177.69	.0000	
13+	0.18	0.01	0.71	1	293.09	.0000	
Age × Race-ethnicity				3	10.44	.0000	0.005430
Non-Hispanic white	0.21	0.01	2.05	1	212.28	.0000	
Non-Hispanic black	0.38	0.03	1.06	1	211.41	.0000	
Mexican American	0.32	0.03	0.63	1	111.51	.0000	
Other	0.18	0.04	1.94	1	18.49	.0001	
Last grade attended × Race-ethnicity				9	4.68	.0002	0.006929
Family income × Race-ethnicity				9	4.34	.0003	0.004385
Sex × Last grade attended × Race-ethnicity				9	2.64	.0142	0.002238
Last night's sleep				2	1.62	.2079	0.000651
Energy level				4	0.94	.4496	0.000584
Computer or video game familiarity				2	64.36	.0000	0.022002
Drinks with caffeine in last 3 h				2	2.48	.0942	0.000442
Drinks with alcohol in last 3 h				2	0.48	.6191	0.000065
Preferred hand				1	2.93	.0934	0.000575
Effort				3	7.27	.0004	0.003558
Test language				1	27.11	.0000	0.008233
Test room temperature	0.17	0.05	1.02	1	10.37	.0023	0.001504
Phase	−0.71	0.25	2.15	1	7.90	.0071	0.002422

Denominator *df* = 49.
R^2 = .482700.
b: slope.
S.E.: standard error.
DEFF: design effect.
SSPCC: squared semipartial correlation coefficient.

Table 11
Summary table for number of errors

Variable	*b*	S.E.	DEFF	*df*	F	P	SSPCC
Sex				1	9.90	.0028	0.004976
Age	0.03	0.01	2.70	1	34.74	.0000	0.022467
Last grade attended				3	3.70	.0176	0.006947
Family income				3	5.11	.0037	0.006482
Race-ethnicity				3	6.68	.0007	0.005566
Sex × Family income				3	3.95	.0133	0.005571
Sex × Race-ethnicity				3	0.44	.7279	0.000680
Last grade attended × Race-ethnicity				9	5.89	.0000	0.008964
Family income × Race-ethnicity				9	2.29	.0309	0.011087
Sex × Family income × Race-ethnicity				9	2.87	.0085	0.006100
Last night's sleep				2	1.84	.1703	0.001259
Energy level				4	2.46	.0573	0.007861
Computer or video game familiarity				2	4.22	.0204	0.005613
Drinks with caffeine in last 3 h				2	0.14	.8686	0.000073
Drinks with alcohol in last 3 h				2	4.39	.0176	0.001281
Preferred hand				1	0.15	.7009	0.000075
Effort				3	0.65	.5883	0.000809
Test language				1	0.78	.3810	0.000280
Test room temperature	− 0.02	0.02	1.39	1	0.96	.3329	0.000319
Phase	− 0.27	0.10	2.14	1	7.60	.0082	0.003903

Denominator *df* = 49.
R^2 = .130066.
b: slope.
S.E.: standard error.
DEFF: design effect.
SSPCC: squared semipartial correlation coefficient.

person was being tested. The main light in the room was fluorescent. There was a small incandescent light near a VAX computer terminal located in the room. The temperature in the test room ranged from 15.6 to 33.3 °C.

2.7. Test procedure

The three neurobehavioral tests were administered in English or Spanish at the preference of the subject. After a subject entered the test room and was oriented to the test procedure and equipment, the technician administered the pretest questionnaire, which consisted of six questions. The tests were then automatically run in the sequence: simple reaction time, symbol–digit substitution and serial digit learning. The subject's responses were recorded directly online throughout testing. Both general and task-specific instructions were presented automatically on the screen. In addition, the technician read each set of instructions aloud to the subject exactly as they were presented on the screen. The technician assisted the subject if it appeared that the directions were not understood, or in the event that the computer detected inappropriate test responses, or computer problems or other difficulties occurred. At the end of the serial digit learning test, a message was displayed on the screen to indicate that the testing was completed. The technician administered a brief posttest questionnaire and recorded the ambient temperature of the room.

Two changes in procedure that occurred during the course of the survey should be noted. During Phase I, if subjects forgot their reading glasses, it was noted on the log. Whether or not the subject completed the test without their glasses was left to the judgment of the subject. Some had a minor near vision correction and were able to complete the test, but some needed their glasses to read the instructions and dropped out. During Phase II, four pairs of reading glasses (+1.0, +1.5, +2.0 and +2.5 diopters) were provided for the subjects to choose from.

Table 12
Performance on the symbol–digit substitution test by race-ethnicity and sex

		Mean total latency			Number of errors		
Race-ethnicity sex	*n*	Mean	S.E.	DEFF	Mean	S.E.	DEFF
Non-Hispanic white							
Male	818	22.48	0.27	2.06	0.88	0.06	1.02
Female	994	20.95	0.22	2.09	1.19	0.09	2.03
Non-Hispanic black							
Male	675	26.48	0.50	1.67	1.45	0.09	0.88
Female	915	25.26	0.46	1.57	1.94	0.11	1.36
Mexican American							
Male	741	27.45	0.41	1.30	1.56	0.13	1.96
Female	764	27.89	0.68	2.03	2.00	0.17	2.59
Other							
Male	82	26.48	1.13	0.97	1.24	0.23	1.53
Female	103	25.05	1.08	1.44	1.57	0.39	1.90

Mean total latency is in seconds.
S.E.: standard error.
DEFF: design effect.

Table 13
Performance on the symbol–digit substitution test by race-ethnicity and last grade attended

Race-ethnicity last grade attended	*n*	Mean total latency			Number of errors		
		Mean	S.E.	DEFF	Mean	S.E.	DEFF
Non-Hispanic white							
0–8	85	29.37	0.95	1.05	2.27	0.36	1.06
9–11	221	24.71	0.42	1.40	1.12	0.13	0.85
12	658	22.51	0.26	1.50	1.25	0.11	1.58
13+	841	19.84	0.14	1.59	0.77	0.06	1.46
Non-Hispanic black							
0–8	94	39.46	1.91	0.93	3.14	0.40	0.97
9–11	317	31.06	0.84	1.24	2.31	0.19	1.07
12	660	25.12	0.39	1.53	1.67	0.12	1.56
13+	511	21.69	0.29	1.24	1.21	0.09	1.05
Mexican American							
0–8	533	34.79	0.68	1.33	1.98	0.13	1.04
9–11	311	27.02	0.88	2.14	2.20	0.17	1.08
12	371	22.86	0.41	1.71	1.37	0.16	1.42
13+	273	21.22	0.39	2.26	1.37	0.23	2.28
Other							
0–8	34	38.16	1.96	1.02	2.22	0.50	1.66
9–11	30	27.88	2.69	1.68	0.81	0.33	1.42
12	37	24.16	1.06	1.72	1.02	0.29	1.33
13+	83	20.87	0.50	1.17	1.51	0.39	1.50

Mean total latency is in seconds.
S.E.: standard error.
DEFF: design effect.

The simple reaction time procedure was also changed. The initial procedure, performed at the first four sites, specified 30 reaction time trials using the preferred hand followed by 20 trials using the nonpreferred hand. The procedure was changed to one of administering all 50 trials with the preferred hand in an effort to reduce the amount of time required for the CNS test.

2.8. Quality assurance

Before beginning testing at a new site, diskette recording heads were cleaned, joystick buttons and keyboard digit keys were inspected for proper operation, and the internal clocks of the computers were reset to the correct local time. Additional system checks were periodically conducted which included checking the alignment of recording heads, checking the speed of disk drives and calibrating timing subroutines.

Daily logs were used to keep track of the participants and testing times, and to record the occurrence of problems and unforeseen situations. The logs and test data were used to generate reports and control charts that, in turn, were also used to identify problems and to provide feedback to the technicians administering the tests. Periodic site visits were conducted during which there were discussions with the technicians, and the technicians were observed administering the tests under field conditions. Each MEC received a site visit about three times per year.

2.9. Calculating summary measures

For the simple reaction time test, the mean reaction time of Trials 11–50 was calculated. Values less than or equal to 50 ms or greater than or equal to 750 ms were considered outliers and were not included in the calculation.

For the symbol–digit substitution test, the mean total latency of the four test trials was calculated. The total latency for each trial did not include the time it took to respond to the first item. It was the total time required to respond to Items 2–9. The number of errors made for the four test trials on Items 2–9 was also calculated.

For the serial digit learning test, the number of trials to reach the criterion of two consecutive trials without a mistake was used. The maximum number of trials was eight. If the subject did not reach the criterion by Trial 8, the number of trials was scored as an 8. The total score was also calculated. This score represents the sum of the error scores for each trial. When a subject's response had fewer than six of the eight digits in the correct position, two points were added to the score; one point was added when either six or seven digits were in the correct position; and zero points were added when all eight digits were reported correctly.

Table 14
Performance on the symbol–digit substitution test by family income and race-ethnicity

Family income race-ethnicity	*n*	Mean total latency			Number of errors		
		Mean	S.E.	DEFF	Mean	S.E.	DEFF
< 10,000							
Non-Hispanic white	130	22.40	0.69	1.70	1.34	0.24	0.99
Non-Hispanic black	320	28.55	0.75	0.87	2.30	0.22	1.49
Mexican American	260	28.98	0.62	0.94	2.16	0.26	2.24
Other	31	28.46	2.84	2.40	2.26	0.75	2.10
10,000–29,999							
Non-Hispanic white	559	22.44	0.26	1.27	1.11	0.11	1.79
Non-Hispanic black	709	25.68	0.48	1.54	1.67	0.09	0.81
Mexican American	722	28.19	0.66	2.27	1.72	0.10	0.98
Other	79	25.58	1.40	2.01	1.05	0.13	0.62
30,000–49,999							
Non-Hispanic white	523	21.67	0.32	2.37	1.03	0.09	1.28
Non-Hispanic black	313	24.47	0.53	1.49	1.54	0.13	1.01
Mexican American	243	26.05	1.03	2.60	1.78	0.20	1.07
Other	42	25.38	1.37	0.90	2.03	0.68	1.46
≥ 50,000							
Non-Hispanic white	516	20.62	0.27	1.78	0.91	0.12	2.39
Non-Hispanic black	135	22.71	0.54	0.82	1.21	0.14	0.76
Mexican American	133	22.01	0.64	1.50	1.39	0.19	0.76
Other	23	20.59	0.73	0.78	0.81	0.23	0.87

Mean total latency is in seconds.
Family income is in dollars.
S.E.: standard error.
DEFF: design effect.

Table 15
Performance on the symbol–digit substitution test by race-ethnicity, sex and last grade attended

Race-ethnicity sex last grade attended		*n*	Mean total latency			Number of errors		
			Mean	S.E.	DEFF	Mean	S.E.	DEFF
Non-Hispanic white								
Male	0–8	38	29.98	0.99	0.63	2.07	0.55	0.99
	9–11	108	25.41	0.72	1.97	0.93	0.10	0.64
	12	279	23.77	0.42	1.40	1.07	0.11	0.91
	13+	392	20.38	0.13	0.58	0.67	0.06	1.02
Female	0–8	47	28.89	1.75	1.72	2.44	0.44	0.90
	9–11	113	23.89	0.44	0.84	1.35	0.22	0.76
	12	379	21.37	0.27	1.14	1.41	0.14	1.36
	13+	449	19.30	0.20	1.95	0.89	0.10	1.76
Non-Hispanic black								
Male	0–8	41	40.66	3.94	1.61	2.61	0.74	1.29
	9–11	145	30.50	0.97	1.21	1.90	0.21	0.76
	12	274	25.74	0.51	1.17	1.34	0.13	0.96
	13+	210	22.41	0.40	1.00	1.07	0.13	0.99
Female	0–8	53	38.46	2.58	1.01	3.57	0.57	1.31
	9–11	172	31.62	1.50	1.57	2.72	0.27	1.05
	12	386	24.65	0.61	2.07	1.93	0.19	1.84
	13+	301	21.16	0.35	1.11	1.30	0.13	1.12
Mexican American								
Male	0–8	282	33.81	0.85	1.46	1.79	0.18	1.14
	9–11	149	25.88	0.62	1.42	1.86	0.22	1.03
	12	161	23.18	0.51	1.90	0.91	0.12	0.94
	13+	139	21.85	0.54	1.96	1.52	0.38	2.13
Female	0–8	251	36.00	0.72	0.55	2.23	0.21	1.22
	9–11	162	28.31	1.39	1.61	2.59	0.31	1.57
	12	210	22.55	0.51	1.16	1.79	0.27	1.62
	13+	134	20.47	0.36	1.13	1.19	0.15	1.13
Other								
Male	0–8	16	38.68	2.16	0.63	1.78	0.54	1.34
	9–11	15	28.46	4.87	2.01	0.95	0.56	1.82
	12	14	24.84	0.66	0.79	0.88	0.43	1.63
	13+	37	21.08	0.90	1.45	1.26	0.37	1.67
Female	0–8	18	37.55	3.74	1.69	2.73	0.81	1.79
	9–11	15	27.28	2.28	0.91	0.67	0.39	1.07
	12	23	23.75	1.63	1.76	1.11	0.39	1.17
	13+	6	20.67	0.58	1.08	1.74	0.71	1.73

Mean total latency is in seconds.
S.E.: standard error.
DEFF: design effect.

2.10. Statistical analysis

Estimates were calculated using the CNS sample weights (the variable name for the weights was WTPFCNS6) so that they would be accurate. The weights take into account the probability of selection, noncoverage and nonresponse. The estimates were calculated with the computer program SUDAAN (Release 7.5.4A, Research Triangle Institute, Research Triangle Park, NC). The complex sample design that was used in the survey affects the variances of the estimates. The effect is called a design effect and is represented quantitatively as the ratio of the variance from a complex sample to the variance from a simple random sample. The variances were estimated by SUDAAN using a linear approximation method. Estimates of the design effects have been included in the tables so that standard deviations for simple random samples can be calculated. (Divide the standard error by the square root of the design effect and multiply by the square root of the sample size.)

The demographic variables used in the analysis were sex (HSSEX), age (HSAGEIR), education (last grade attended, HFA8R), family income (HFF19R) and race-ethnicity (DMARETHN). Three variables (CNPQ07, CNPQ08 and CNPQ09) that indicated whether or not a subject completed a test were used to calculate the non-response rates.

The questionnaire variables used were: How much sleep did you get last night? Would you say about the ususal amount, less than usual or more than usual? (CNPQ01), Right now would you say you are feeling energetic, fresh, average, tired or exhausted? (CNPQ02), How much familiarity do you have with computers or video games? Would you say none, some or a lot? (CNPQ03), Not counting decaffeinated products, how many cups of coffee, tea or cola have you had in the past 3 h? (CNPQ04), How many drinks of alcohol have you had in the past 3 h? Include any alcoholic drinks such as wine, beer or hard liquor. (CNPQ05), Which hand do you prefer to use for this test? (CNPQ06). The language the test was administered in (CNPLANG) and the test room temperature (CNPTEMP) were also included as "questionnaire" variables.

Linear models were used to test hypotheses, calculate slopes and to provide a guide for using the demographic and questionnaire variables as covariates. The models were created in a stepwise fashion. In the first step, all main

Table 16
Performance on the symbol–digit substitution test by family income and sex

Family income sex	*n*	Mean total latency			Number of errors		
		Mean	S.E.	DEFF	Mean	S.E.	DEFF
< 10,000							
Male	292	26.03	0.83	2.05	1.15	0.15	1.56
Female	449	24.53	0.67	1.96	2.15	0.26	2.74
10,000–29,999							
Male	953	24.38	0.44	3.20	1.12	0.09	1.93
Female	1116	22.97	0.30	1.72	1.35	0.12	3.03
30,000–49,999							
Male	536	22.97	0.40	2.42	0.96	0.08	1.02
Female	585	21.57	0.32	1.87	1.37	0.12	1.69
≥ 50,000							
Male	375	21.73	0.36	1.61	0.87	0.12	1.87
Female	432	19.81	0.22	1.54	0.98	0.11	1.89

Mean total latency is in seconds.
Family income is in dollars.
S.E.: standard error.
DEFF: design effect.

Table 17
Performance on the symbol–digit substitution test by family income, race-ethnicity and sex

Family income race-ethnicity	sex	*n*	Mean total latency			Number of errors		
			Mean	S.E.	DEFF	Mean	S.E.	DEFF
< 10,000								
Non-Hispanic white	Male	45	23.47	0.91	0.90	0.82	0.18	0.74
	Female	85	21.76	0.79	1.58	1.65	0.40	1.25
Non-Hispanic black	Male	116	29.61	1.41	1.31	1.59	0.25	1.12
	Female	204	27.90	1.04	0.97	2.74	0.29	1.49
Mexican American	Male	115	28.84	0.93	1.09	1.98	0.44	2.67
	Female	145	29.11	0.94	1.06	2.32	0.23	1.05
Other	Male	16	27.21	4.35	2.13	0.88	0.36	0.97
	Female	15	30.48	2.22	1.66	4.47	0.95	1.26
10,000–29,999								
Non-Hispanic white	Male	262	23.20	0.36	1.08	0.98	0.11	1.05
	Female	297	21.65	0.28	0.87	1.23	0.17	1.81
Non-Hispanic black	Male	303	26.31	0.61	1.11	1.43	0.15	1.00
	Female	406	25.15	0.74	1.99	1.86	0.13	0.76
Mexican American	Male	358	28.09	0.82	2.10	1.53	0.14	0.98
	Female	364	28.31	0.85	1.62	1.92	0.16	1.19
Other	Male	30	27.46	2.59	1.84	1.28	0.25	0.89
	Female	49	24.16	1.06	1.11	0.88	0.22	1.14
30,000–49,999								
Non-Hispanic white	Male	244	22.44	0.46	2.19	0.83	0.09	0.72
	Female	279	20.87	0.31	1.23	1.23	0.12	0.99
Non-Hispanic black	Male	150	24.68	0.62	0.95	1.63	0.27	1.63
	Female	163	24.27	0.68	1.28	1.47	0.18	1.15
Mexican American	Male	120	26.84	0.96	1.46	1.42	0.20	0.94
	Female	123	25.11	1.41	1.92	2.20	0.37	1.22
Other	Male	22	25.18	2.23	0.93	1.44	0.48	1.21
	Female	20	25.65	1.63	1.01	2.81	1.43	1.63
≥ 50,000								
Non-Hispanic white	Male	231	21.55	0.39	1.23	0.86	0.14	1.40
	Female	285	19.73	0.23	1.29	0.95	0.13	1.74
Non-Hispanic Black	Male	61	25.20	0.83	0.59	0.86	0.20	1.13
	Female	74	20.73	0.61	1.35	1.49	0.24	0.92
Mexican American	Male	74	22.21	0.70	1.35	1.33	0.33	1.33
	Female	59	21.68	0.89	0.88	1.50	0.40	1.28
Other	Male	9	22.34	1.46	0.90	0.81	0.46	1.16
	Female	14	19.72	0.99	1.18	0.82	0.32	1.11

Mean total latency is in seconds.
Family income is in dollars.
S.E.: standard error.
DEFF: design effect.

effects were entered into the model. In the second step, two-way interactions between demographic factors that had a statistically significant ($P < .05$) main effect were entered. The two-way interactions were dropped if they were not statistically significant and they were not contained in a three-way interaction to be tested in the next step. In the third step, three-way interactions between demographic variables were entered by combining statistically significant two-way interactions with a common variable. Three-way interactions were dropped if they were not statistically significant or if they had empty cells. Two-way interactions were dropped if they were not statistically significant and were not contained in a three-way interaction remaining in the model. No demographic variable by questionnaire variable interactions were tested. Main effects were always retained in the models.

Test room temperature was recorded in degrees Fahrenheit, but analyzed and reported in degrees Celsius. Age and test room temperature were treated as continuous variables in the models. Slope coefficients for these variables appear in the summary tables that follow. A variable called "phase" was included in the models to account for the change in monitors and the change in procedure with regard to glasses between the two phases of the survey. The slope coefficients for this variable represent performance differences between Phases I and II. The mean reaction times from persons who used their preferred and nonpreferred hands were not included in the analyses.

3. Results

Table 2 shows the proportion of persons who did not complete the tests by demographic categories. The nonresponse rates increased from test to test in the order that the tests were administered. The nonresponse rates of the males were higher than the rates of the females. As age increased, the nonresponse rates increased. The nonresponse rates tended to decrease as the level of education increased. The nonresponse rates of the non-Hispanic white group were lower than those of the other racial-ethnic groups. Reasons for nonresponse included equipment malfunction, insufficient time to administer the tests, the test room was not available, the subject refused or was uncooperative, the subject was physically unable to perform the tests, the subject did not understand the test instructions due to a language barrier or other reasons, and the subject became excessively frustrated, angry or inattentive.

The race-ethnicity variable was based on reported race and ethnicity. All blacks and whites of non-Hispanic descent took the tests in English. Of the Mexican Americans who took the tests, 41% took them in Spanish. Hispanics who were not Mexican–American were included in the Other category, along with non-Hispanics who did not say they were white or black. Of those in the Other category who took the tests, 31% took them in Spanish.

Table 18
Performance on the serial digit learning test by demographic categories

Demographic variable	*n*	Trials to criterion			Total score		
		Mean	S.E.	DEFF	Mean	S.E.	DEFF
Overall	4962	4.65	0.07	5.48	4.48	0.15	6.03
Sex							
Male	2258	4.60	0.09	3.98	4.44	0.18	3.69
Female	2704	4.70	0.07	3.50	4.52	0.18	4.25
Age							
20–29	1592	4.29	0.09	3.04	3.68	0.19	3.69
30–39	1433	4.47	0.10	3.60	4.16	0.21	3.57
40–49	1145	4.85	0.11	2.96	4.83	0.22	2.79
50–59	792	5.36	0.14	3.10	6.02	0.30	2.83
Last grade attended							
0–8	676	6.79	0.14	4.28	9.98	0.38	3.95
9–11	841	5.54	0.16	4.72	6.41	0.31	3.51
12	1713	5.00	0.07	2.16	4.98	0.16	2.17
13+	1701	3.83	0.06	2.38	2.79	0.12	2.59
Family income							
<10,000	707	5.15	0.15	3.07	5.84	0.38	3.78
10,000–29,999	2005	5.06	0.08	2.92	5.37	0.19	3.15
30,000–49,999	1110	4.53	0.08	1.79	4.07	0.17	2.14
≥50,000	797	3.96	0.10	2.68	3.04	0.19	2.44
Race-ethnicity							
Non-Hispanic white	1801	4.36	0.07	2.54	3.79	0.16	2.99
Non-Hispanic black	1549	5.38	0.06	1.30	6.14	0.16	1.61
Mexican American	1433	6.12	0.11	4.45	7.98	0.31	5.35
Other	179	5.60	0.20	1.52	6.78	0.52	1.80

Age is in years.
Family income is in dollars.
S.E.: standard error.
DEFF: design effect.

Table 19
Performance on the serial digit learning test by questionnaire categories

Questionnaire variable	*n*	Trials to criterion			Total score		
		Mean	S.E.	DEFF	Mean	S.E.	DEFF
Last night's sleep							
Less than usual	1482	4.68	0.11	4.28	4.67	0.24	4.12
About usual amount	2974	4.70	0.07	3.36	4.49	0.15	3.82
More than usual	446	4.26	0.12	1.51	3.87	0.31	2.09
Energy level							
Energetic	270	4.68	0.21	2.77	4.57	0.41	2.51
Fresh	635	4.60	0.12	2.17	4.35	0.26	2.41
Average	2693	4.65	0.07	3.47	4.40	0.16	3.88
Tired	1169	4.63	0.12	3.65	4.56	0.23	2.98
Exhausted	136	5.08	0.31	2.29	6.11	0.71	1.97
Computer or video game familiarity							
None	1986	5.75	0.08	2.98	7.03	0.20	2.97
Some	2220	4.33	0.08	3.76	3.69	0.16	3.95
A lot	697	3.79	0.10	2.06	2.64	0.16	1.87
Drinks with caffeine in last 3 h							
0	4689	4.64	0.07	5.62	4.46	0.16	5.85
1	151	4.93	0.29	2.22	5.44	0.68	2.30
2+	63	4.57	0.39	2.07	4.39	0.85	2.18
Drinks with alcohol in last 3 h							
0	4855	4.65	0.07	5.57	4.47	0.16	6.04
1	37	6.85	0.23	0.81	9.58	0.84	1.13
2+	11	4.72	0.87	1.88	4.53	1.44	1.14
Preferred hand							
Right	4500	4.67	0.08	5.99	4.53	0.17	6.44
Left	403	4.48	0.12	1.58	3.95	0.22	1.32
Effort							
None	107	4.11	0.27	1.95	3.61	0.54	1.67
Some	1210	4.52	0.12	4.02	4.37	0.27	4.20
A lot	1100	4.67	0.09	2.20	4.30	0.19	2.47
As hard as I could	2485	4.71	0.08	3.70	4.63	0.18	4.13
Test language							
English	4284	4.55	0.07	4.68	4.23	0.15	5.26
Spanish	619	7.12	0.14	5.86	10.70	0.42	5.28
Test room temperature							
15.6–19.4	30	4.04	0.49	1.77	3.44	0.81	1.26
20.0–21.1	278	4.12	0.21	3.10	3.49	0.39	2.69
21.7–22.8	972	4.43	0.12	3.55	4.13	0.26	3.22
23.3–24.4	1571	4.64	0.09	3.04	4.34	0.20	3.29
25.0–26.1	855	5.17	0.13	3.02	5.57	0.28	3.06
26.7–27.8	164	5.19	0.23	1.89	5.62	0.40	1.19
28.3–33.3	9	7.63	0.25	0.45	10.80	0.31	0.13

Test room temperature is in degrees Celsius.
S.E.: standard error.
DEFF: design effect.

Table 20
Summary table for trials to criterion

Variable	*b*	S.E.	DEFF	*df*	F	P	SSPCC
Sex				1	4.78	.0336	0.001675
Age	0.04	0.00	1.79	1	73.13	.0000	0.025912
Last grade attended				3	67.86	.0000	0.049422
Family income				3	8.68	.0001	0.007334
Race-ethnicity				3	39.14	.0000	0.019557
Last grade attended × Family income				9	2.39	.0246	0.004340
Family income × Race-ethnicity				9	3.33	.0030	0.003349
Last night's sleep				2	3.94	.0259	0.001459
Energy level				4	0.55	.7003	0.000757
Computer or video game familiarity				2	19.64	.0000	0.013092
Drinks with caffeine in last 3 h				2	0.17	.8407	0.000161
Drinks with alcohol in last 3 h				2	4.52	.0158	0.000524
Preferred hand				1	0.96	.3320	0.000419
Effort				3	6.76	.0007	0.007827
Test language				1	17.58	.0001	0.004478
Test room temperature	0.12	0.03	2.67	1	14.33	.0004	0.007590
Phase	− 0.09	0.12	3.54	1	0.56	.4595	0.000390

Denominator *df* = 49.
R^2=.292050.
b: slope.
S.E.: standard error.
DEFF: design effect.
SSPCC: squared semipartial correlation coefficient.

Table 3 shows performance on the tests at selected percentiles. All the values for the continuous variables were able to be estimated. Some of the values for the discrete variables could not be estimated as indicated by

Table 21
Summary table for total score

Variable	*b*	S.E.	DEFF	*df*	F	P	SSPCC
Sex				1	0.24	.6259	0.000079
Age	0.07	0.01	1.59	1	82.03	.0000	0.024242
Last grade attended				3	68.46	.0000	0.047207
Family income				3	9.31	.0001	0.006982
Race-ethnicity				3	39.37	.0000	0.019849
Age × Family income				3	1.44	.2436	0.002157
Age × Race-ethnicity				3	4.28	.0092	0.001311
Non-Hispanic white	0.06	0.01	1.84	1	40.45	.0000	
Non-Hispanic black	0.12	0.01	0.50	1	81.54	.0000	
Mexican American	0.11	0.02	0.34	1	46.65	.0000	
Other	0.08	0.03	2.75	1	5.29	.0258	
Last grade attended × Family income				9	2.70	.0124	0.004832
Family income × Race-ethnicity				9	3.65	.0015	0.003069
Age × Family income × Race-ethnicity				9	3.84	.0010	0.002301
Last night's sleep				2	0.87	.4263	0.000498
Energy level				4	0.33	.8595	0.000439
Computer or video game familiarity				2	33.50	.0000	0.016746
Drinks with caffeine in last 3 h				2	0.48	.6219	0.000146
Drinks with alcohol in last 3 h				2	4.69	.0136	0.000723
Preferred hand				1	1.47	.2317	0.000536
Effort				3	4.62	.0064	0.003751
Test language				1	46.79	.0000	0.006910
Test room temperature	0.21	0.06	2.28	1	12.39	.0009	0.005327
Phase	− 0.26	0.24	3.45	1	1.22	.2742	0.000800

Denominator *df* = 49.
R^2 = .325346.
b: slope.
S.E.: standard error.
DEFF: design effect.
SSPCC: squared semipartial correlation coefficient.

the blank spaces. There were not enough unique values of the discrete variables to do the interpolation necessary to calculate the percentiles.

3.1. Simple reaction time

Performance on the simple reaction time test is shown by demographic categories in Table 4 and by questionnaire categories in Table 5. A summary of the analysis with a linear model is shown in Table 6.

Males had a lower average mean reaction time than females. The average mean reaction time increased as age increased, and decreased as education level and family income increased. The non-Hispanic white group had a lower average mean reaction time than the non-Hispanic black group, the Mexican American group and the Other group.

The average mean reaction time was directly related to the reported amount of last night's sleep and test room temperature, and was inversely related to the reported energy level.

The Family income × Race-ethnicity interaction was statistically significant. The differences in average mean reaction time between race-ethnicity categories tended to decrease as family income increased (Table 7).

3.2. Symbol–digit substitution

Performance on the symbol–digit substitution test is shown by demographic categories in Table 8 and by questionnaire categories in Table 9. Summaries of the analyses with linear models are shown in Tables 10 and 11.

3.2.1. Mean total latency

Males had a higher average mean total latency than females. The average mean total latency increased as age increased, and decreased as education level and family income increased. The non-Hispanic white group had a lower average mean total latency than the non-Hispanic black group, the Mexican American group and the Other group.

The average mean total latency was related to the reported amount of familiarity with computers or video games and the reported amount of effort. Those who took the test in Spanish had a higher average mean total latency than those who took the test in English. The average mean total latency increased as test room temperature increased, and decreased from Phases I to II.

The Sex × Race-ethnicity (Table 12), Age × Education level (Table 10), Age × Race-ethnicity (Table 10), Education level × Race-ethnicity (Table 13), Family income × Race-ethnicity (Table 14) and Sex × Education level × Race-ethnicity (Table 15) interactions were statistically significant. The slope between mean total latency and age decreased as education level increased, and race-ethnicity differences in average mean total latency tended to decrease as family income increased.

3.2.2. Number of errors

Males had a lower average number of errors than females. The average number of errors increased as age increased, and decreased as education level and family income increased. The non-Hispanic white group had a lower average number of errors than the non-Hispanic black group, the Mexican American group and the Other group.

The average number of total errors was related to familiarity with computers or video games and the reported number of alcoholic beverages in the last 3 h. The average number of total errors decreased from Phases I to II.

The Sex × Family income (Table 16), Education level × Race-ethnicity (Table 13), Family income × Race-ethnicity (Table 14) and Sex × Family income × Race-ethnicity (Table 17) interactions were statistically significant. Sex and race-ethnicity differences in the average number of errors tended to decrease as family income increased.

3.3. Serial digit learning

Performance on the serial digit learning test is shown by demographic categories in Table 18 and by questionnaire categories in Table 19. Summaries of the analyses with linear models are shown in Tables 20 and 21.

Table 22
Performance on the serial digit learning test by family income and last grade attended

Family income last grade attended	*n*	Trials to criterion			Total score		
		Mean	S.E.	DEFF	Mean	S.E.	DEFF
< 10,000							
0–8	158	7.15	0.17	2.69	10.64	0.57	2.83
9–11	185	5.75	0.27	2.88	6.99	0.76	4.06
12	225	5.30	0.22	2.14	6.13	0.55	2.72
13+	134	3.59	0.13	1.23	2.30	0.21	0.94
10,000–29,999							
0–8	350	6.88	0.15	2.69	10.04	0.45	2.91
9–11	425	5.56	0.19	3.51	6.47	0.37	2.41
12	685	5.23	0.12	2.12	5.51	0.28	2.51
13+	535	4.06	0.11	1.65	3.23	0.19	1.56
30,000–49,999							
0–8	65	5.58	0.42	2.71	7.17	0.97	2.28
9–11	138	5.58	0.37	3.77	6.32	0.73	3.22
12	460	4.90	0.12	1.76	4.55	0.22	1.49
13+	442	3.89	0.12	2.15	2.89	0.23	2.25
≥ 50,000							
0–8	15	4.83	0.94	2.28	5.71	2.02	2.31
9–11	33	4.98	0.33	1.11	5.60	0.80	1.09
12	234	4.52	0.20	2.17	4.09	0.41	2.18
13+	511	3.67	0.10	1.99	2.45	0.16	1.91

Family income is in dollars.
S.E.: standard error.
DEFF: design effect.

3.3.1. Trials to criterion

Males had a lower average number of trials to criterion than females. The average number of trials to criterion increased as age increased, and decreased as education level and family income increased. The non-Hispanic white group had a lower average number of trials to criterion than the non-Hispanic black group, the Mexican American group and the Other group.

The average number of trials to criterion was related to the reported amount of last night's sleep, the reported amount of familiarity with computers or video games, the reported number of alcoholic beverages in the last 3 h, and the reported amount of effort. Those who took the test in Spanish had a higher average number of trials to criterion than those who took the test in English. As the test room temperature increased, the average number of trials to criterion increased.

The Education level × Family income (Table 22) and the Family income × Race-ethnicity (Table 23) interactions were statistically significant. Education level and race-ethnicity differences in the average number of trials to criterion tended to decrease as family income increased.

3.3.2. Total score

The average total score increased as age increased, and decreased as education level and family income increased. The non-Hispanic white group had a lower average total score than the non-Hispanic black group, the Mexican American group and the Other group.

The average total score was related to the reported amount of familiarity with computers or video games, the reported number of alcoholic beverages in the last 3 h and the reported amount of effort. Those who took the test in Spanish had a higher average total score than those who took the test in English. As the test room temperature increased, the average total score increased.

The Age × Race-ethnicity (Table 21), Education level × Family income (Table 22), Family income × Race-ethnicity (Table 23) and Age × Family income × Race-ethnicity (Table 24) interactions were all statistically significant. Education level and race-ethnicity differences in the average total score tended to decrease as family income increased.

Table 23
Performance on the serial digit learning test by family income and race-ethnicity

Family income race-ethnicity	*n*	Trials to criterion			Total score		
		Mean	S.E.	DEFF	Mean	S.E.	DEFF
< 10,000							
Non-Hispanic white	129	4.59	0.24	1.66	4.59	0.63	2.25
Non-hispanic black	306	5.71	0.12	1.01	7.06	0.32	1.18
Mexican American	242	6.57	0.16	1.94	9.00	0.45	2.12
Other	30	5.66	0.68	2.37	7.03	1.58	2.35
10,000–29,999							
Non-Hispanic white	555	4.71	0.11	1.39	4.49	0.23	1.59
Non-Hispanic black	690	5.39	0.08	0.90	6.19	0.23	1.38
Mexican American	683	6.26	0.13	2.86	8.36	0.39	3.93
Other	77	6.13	0.24	1.07	8.09	0.72	1.33
30,000–49,999							
Non-Hispanic white	523	4.32	0.10	1.32	3.61	0.20	1.66
Non-Hispanic black	311	5.33	0.16	1.81	5.78	0.38	2.05
Mexican American	237	5.89	0.19	1.97	7.28	0.44	1.67
Other	39	5.51	0.28	0.77	6.00	0.65	0.97
≥ 50,000							
Non-Hispanic white	510	3.90	0.11	2.04	2.93	0.20	1.94
Non-Hispanic black	134	4.77	0.20	1.14	4.58	0.40	1.08
Mexican American	130	4.97	0.22	1.65	4.98	0.42	1.46
Other	23	3.87	0.20	0.38	2.92	0.35	0.38

Family income is in dollars.
S.E.: standard error.
DEFF: design effect.

Table 24
Slopes between total score and age by family income and race-ethnicity

Family income race-ethnicity	*b*	S.E.	DEFF	*df*	t	P
< 10,000						
Non-Hispanic white	0.08	0.04	2.54	1	2.16	.0361
Non-Hispanic black	0.12	0.02	0.38	1	5.21	.0000
Mexican American	0.12	0.03	0.22	1	3.77	.0004
Other	−0.00	0.07	1.06	1	−0.01	.9896
10,000–29,999						
Non-Hispanic white	0.10	0.02	3.40	1	4.40	.0001
Non-Hispanic black	0.11	0.03	0.97	1	4.11	.0002
Mexican American	0.13	0.02	0.41	1	5.18	.0000
Other	0.08	0.06	4.41	1	1.26	.2121
30,000–49,999						
Non-Hispanic white	0.03	0.02	1.85	1	1.64	.1071
Non-Hispanic Black	0.11	0.04	1.03	1	2.73	.0089
Mexican American	0.12	0.02	0.10	1	5.85	.0000
Other	0.14	0.06	2.19	1	2.43	.0189
≥ 50,000						
Non-Hispanic white	0.05	0.01	1.01	1	3.70	.0005
Non-Hispanic black	0.15	0.03	0.24	1	5.71	.0000
Mexican American	0.02	0.04	0.25	1	0.42	.6740
Other	0.01	0.08	2.07	1	0.17	.8629

Family income is in dollars.
Denominator *df* = 49.
b: slope.
S.E.: standard error.
DEFF: design effect.

4. Discussion

All of the demographic variables that were considered were related to neurobehavioral test performance. The performance of males and females differed on all the tests. Performance on all the tests decreased as age increased. Performance on all the tests increased as education level and family income increased. There were also consistent race-ethnicity differences.

For the questionnaire variables, persons who reported having more sleep performed better on the simple reaction time and serial digit learning tests. Persons who reported feeling less energetic performed worse on the simple reaction time test. As computer or video game familiarity increased, performance on the symbol–digit substitution and serial digit learning tests increased. Persons who reported having one alcoholic beverage within the last 3 h did not perform as well as those who reported otherwise on the symbol–digit substitution and serial digit learning tests. Reported effort and the language the test was given in were related to performance on the symbol–digit substitution and the serial digit learning tests. Performance on all the tests decreased as the temperature of the test room increased.

Family income interacted with sex, education level and race-ethnicity. As family income increased, the difference in performance between sexes, racial-ethnic groups and levels of education tended to decrease. The slope between mean total latency from the symbol–digit substitution test and age decreased as education level increased. The slopes between age and mean total latency from the symbol–digit substitution test, and age and total score from the serial digit learning test varied by racial-ethnic group.

4.1. Using covariates

All the variables discussed in the previous three paragraphs should be considered when designing a study, developing study procedures or analyzing data. Not all variables will be important for all studies. Interactions between covariates should be kept in mind, although for purposes of adjustment, they probably do not have to be included in statistical models. The need for unbiased estimates of exposure effects must be balanced with the principle of parsimony.

An exposure effect may manifest itself as an interaction with a covariate. Krieg et al. [31] found statistically significant interactions between age and exposure to organophosphate pesticides on several measures of simple reaction time performance. An Age × Exposure interaction may indicate a longer exposure if no variable is included to account for duration of exposure, or it may mean that older workers are more susceptible to the effects of exposure than younger workers.

If an exposure affects a covariate, then including the covariate in a statistical model may reduce or eliminate the effect of the exposure. For example, vocabulary test performance is related to neurobehavioral test performance, and vocabulary tests are often used to control for intellectual ability [41] even though there is evidence that vocabulary test performance is related to solvent [13,27], lead [5] and mercury vapor [23] exposures. Similarly, short-term exposures to organic solvents are known to produce fatigue and to interfere with sleep [9]. A variable that is affected by a toxic exposure should not be used as a covariate to adjust the exposure effect. It may be possible to capture the direct effect of an exposure on performance as well as an indirect effect through another variable by using a structural equation model.

4.2. Previous NES studies

Previous studies which report the relationships between performance on the NES tests and demographic and other covariates fall into two categories: methodological studies which investigate covariates [4,14,29,37,41,43,52,54] and studies of exposure effects which include additional information about covariates [2,19,24,47,48,65]. Table 25 summarizes the results of these studies. Each symbol in the table represents the results of one analysis of one set of data. In several instances, the same data are analyzed twice in different ways. Not all of the variables used in the previous studies are included in the table. For sex, a plus sign indicates that the value for the females is greater than the males.

Age and symbol–digit substitution latency, and age and serial digit learning total score are consistently related in the literature. These relationships are confirmed here. Letz et al.

Table 25
Results of previous NES studies

	Simple reaction time	Symbol–digit substitution		Serial digit learning	
Variable	Mean reaction time	Mean total latency	Number of errors	Trials to criterion	Total score
Sex	–	– –			
	ØØ*	ØØØ	Ø		Ø
	++		+		
Age	–				
	ØØØØØ*	Ø*****	Ø		***
	+	++++	+		+
Education		– –	–		
	ØØØØØØ**	ØØØ***			ØØØ*
Race	Ø*	*			Ø*
Computer or video game experience	– Ø	Ø*			ØØ
Effort		ØØ			Ø*
Vocabulary test		– –	–		
		***	Ø		**
Technical test		–			–
	Ø				
Income	–	–			–
Alcohol consumption	–				–
	ØØ	ØØØØØØ*			ØØ
	+	+			
Smoking	Ø	Ø			
Day of week		ØØ			ØØ
Time of day		ØØ			Ø*
Contrast sensitivity	– –	– – –	–		
	ØØ		ØØ		
	+				

–: statistically significant inverse relationship.
Ø: no statistically significant relationship detected.
*: statistically significant relationship of unspecified direction.
+: statistically significant direct relationship.

[41] found that income was inversely related to mean reaction time, symbol–digit substitution latency and serial digit learning total score. These relationships are corroborated here. Vocabulary test score and symbol–digit substitution latency, and vocabulary test score and serial digit learning total score are also consistently related in the literature. Contrast sensitivity was consistently related to symbol–digit substitution latency in one study [29]. NHANES III did not include a vocabulary test for adults, or measurements of contrast sensitivity.

Inconsistent or no relationships have been found in the literature between simple reaction time and sex, age, education and race; symbol–digit substitution latency and sex, education, computer or video game experience and effort; symbol–digit substitution errors and sex and age; and serial digit learning total score and education, race, computer or video game experience and effort. All these relationships were statistically significant in the present study. The inconsistent or lack of relationships may be explained by the relatively small sample sizes or the range of values of the covariates used in the previous studies.

The manner in which results are reported in previous studies is inconsistent. Some investigators report slopes or correlation coefficients; some report the direction of an effect, but do not indicate its magnitude; some report that performance was better or worse, but do not give the literal direction of the relationship; some report proportions of variance, but do not explicitly give the direction of the relationship or the magnitude in terms of slopes or means; and some report only statistically significant relationships, while others include relationships that are not statistically significant. This makes it difficult to combine information from different studies. Ideally, regression coefficients or means would be reported, even if a relationship is not statistically significant.

The results with regard to covariates have implications for exposure effects. If an exposure effect has a magnitude similar to the differences between some of the demographic or questionnaire categories in the present study, it may require larger sample sizes or multiple studies to detect the effect. To this end, effect sizes should be reported in their units of measurement along with estimates of precision so that results can be combined. A study that reports no effect should indicate how big of an effect it could detect.

4.3. Temperature and performance

Non-computerized versions of the neurobehavioral tests used in NHANES III have been used to study the effect of temperature on performance. The results for simple reaction time are mixed, some finding a decrease in reaction time, some finding no effect, and some finding an increase. Lovingood et al. [45] found that simple visual reaction time decreased during exposure to an ambient temperature of 125.6°F. Reilly and Parker [57] found that simple visual and auditory reaction times were shorter during exposure to an ambient temperature of 86°F. Holt and Brainard [28] found that simple reaction time decreased when subjects were exposed to heat. Benor and Shvartz [10] found that simple visual reaction time was not affected by temperatures ranging from 30 to 50 °C, and Aird et al. [1] found no effect of a heat exposure of 36 °C on total simple reaction time. Shvartz et al. [62] found an increase in simple reaction time in subjects exposed to heat (40 °C DB, 30 °C WB), and Razmjou and Kjellberg [56] found that simple reaction time increased when a group of subjects was exposed to a temperature of 40 °C. Rogers and Noddin [58] found that exposure to cold temperatures (1 and −10 °C) did not have an effect on simple reaction time.

The results are more consistent for the other two tests, albeit there are fewer studies. Bhattacharya et al. [11] found that subjects did not perform as well on a digit span test when they were in a hot environment (34 °C CET) as compared to a comfortable one (24 °C CET). Bhattacharya et al. [12] found that the average number of digit–symbol attempts decreased as temperature increased (25, 30 and 35 °C), and that in a quiet noise condition, the number of correct attempts decreased and the number of errors increased as temperature increased. Kumar et al. [33] found that workers in a glass bangle factory exposed to an ambient temperature of 38.2 °C performed worse than controls on both a digit–symbol and a digit span test.

In the present study, relationships between temperature and performance have been demonstrated at temperatures less extreme than those studied previously. It is important to control the test room temperature when administering neurobehavioural tests, and to report what the room temperature was during testing.

4.4. Percentiles

Percentiles were calculated and included in Table 2 in order to explore their use as reference values. The continuous variables provide useful estimates, but the count variables, i.e., the number of errors and trials, do not. The count variables do not have enough unique values to provide estimates for many of the percentiles.

If the mean performance of an exposed group is less than a control group, and the mean of the exposed group is in a high percentile and the mean of the control group is not, then this could lend further support to an exposure effect, but it could also mean that the control and exposed groups are not matched properly for demographic variables. If exposed and control groups are matched properly, it is more likely that the group means would have similar percentiles and that the mean of the exposed group is only slightly different from the control group. Still it is worthwhile to compare group means to the percentiles, to know where the groups lie with regard to the population as a whole, and potentially to identify groups with extreme scores. Percentiles should not be used as a replacement for a control group.

Comparing individual performance to the percentiles in Table 2 may indicate nervous system dysfunction, but it may also indicate a demographic difference to some degree. Assuming the comparison is valid, or assuming demographically adjusted tables are available, a high percentile may indicate dysfunction. In order for a neurobehavioral test to be useful for diagnostic purposes or as a screening device for individuals, one would have to correlate the performance on the test to the physical functioning of the nervous system. The same criteria that apply to biochemical tests, sensitivity and specificity, should apply to neurobehavioral tests. Given the amount of variability in performance between individuals, it is unlikely that the neurobehavioral tests considered in this paper can be used to detect preclinical dysfunction in an individual. However, White et al. [67] found that a group of patients with multiple sclerosis and another group that had Parkinson's disease did not perform as well as a control group on many NES tests, indicating that the test battery can be used to identify groups with clinical symptoms. More work needs to be done to correlate neurobehavioral test performance with the symptoms of neurological diseases, and to determine which areas of the nervous system are used and active when neurobehavioral tests are being performed. Such information could extend the functionality of the tests and enhance their interpretability.

4.5. Other test batteries

Other test batteries have tests similar to the simple reaction time and symbol–digit substitution tests found in the NES. There are methodological studies for these batteries that give reference values for the tests they contain and that provide information about the relationships between demographic or other variables and performance on the batteries. Some of these studies are discussed in this section. This discussion is not intended to be inclusive in terms of test batteries or studies using a given test battery.

The Wechsler Adult Intelligence Scale — Revised (WAIS-R) contains a digit symbol test [66]. The norms for the WAIS-R were based on a stratified, quota sample of 1880 persons, 16–74 years old, from the United States. The variables used for stratification were age, sex, race, geographic region, occupation, education and urban or rural residence. Persons were tested from 1976 to 1980. Figures from the 1970 United States Census and some later Census reports were used to stratify the sample.

An analysis of the WAIS-R subtests [30] indicated that digit symbol performance was related to sex, race, geographic region and education, but not to urban or rural residence. Females had higher average scores than males, whites had higher average scores than blacks, and the average scores increased as education increased. Occupation was not analyzed in this study, and a separate analysis was done for each of four age groups. The age groups were not directly compared.

The Pittsburgh Occupational Exposures Test Battery [59] contains a symbol–digit learning test. Norms were calculated for the battery using a random sample of 182 white males from an industrial plant in Pennsylvania. Performance on the symbol–digit test was related to age ($r = -.416$) and education ($r = .326$).

The Milan Automated Neurobehavioral System is a computerized battery that contains simple reaction time and symbol digit tests [18]. In a study of 20 smelters, performance on the simple reaction time test was directly related to age, but was not related to years of school or alcohol intake. Performance on the symbol digit test was not related to age, years of school or alcohol intake.

The Swedish Performance Evaluation System is a computerized battery that contains simple reaction time and symbol digit tests [25]. One standardization study was done using 38 university students and 62 employees of the National Institute of Occupational Health. No differences between sexes, age groups, educational groups or groups with different amounts of computer experience were found for the simple reaction time test. No sex or computer experience differences were found for the symbol digit test, however, the oldest age group (40 and older) had a higher average latency than the two other age groups (30–39, 29 and younger), and as education level increased average latency decreased. A second standardization study was done for simple reaction time using 730 industrial workers. Using regression analysis, no relationships between age or time of day and performance were found.

Williamson [68] developed a test battery that contained a simple reaction time test. Using a sample of 228 nonexposed workers, she found that performance on the test was related to age, but not to education, job type, sex or length of residence in Australia.

A cross-cultural assessment has been done for the World Health Organization Neurobehavioral Core Test Battery [17]. This battery contains simple reaction time and digit symbol tests. Each participating country was to test a primary group of 50 male or female subjects, 25–36 years old. An analysis of the European data indicated that there was an inverse relationship between simple reaction time and years of school completed, and that performance on the simple reaction time and digit symbol tests varied between countries [17]. The Chinese portion of the assessment used 282 workers, 16–55 years old, from two fabric weaving manufacturers in Shanghai [42]. An analysis of this data indicated that the average reaction times of males were less than females, and that the average reaction time tended to increase as age increased. Performance on the digit symbol test tended to decrease with age, and the females tended to perform better than the males. A descriptive analysis of all the data from the assessment indicated that simple reaction time performance was similar between countries, but that performance on the digit symbol test was more variable [3].

A second study using the World Health Organization battery was done with 715 subjects recruited in California,

Georgia and Oregon in the United States [4]. Results from one-way analyses of variance indicated that education level, cultural group, gender, and age were related to simple reaction time and digit symbol performance, but that the size of the city where a person was educated was not related to performance on the tests. The average reaction time of the males was less than the females. With respect to the digit symbol test, as education level increased performance increased, and the females performed better than the males.

The statistically significant results found with the other test batteries that were reviewed here are consistent with the results of the present study. There were relationships found in the present study that were not detected in some of the studies using other test batteries. This may be due to the larger sample size used in the present study.

Attention has been paid to the method of sampling used in the studies. Only one study [59] used a probability sample. The other studies used nonprobability samples. Generalizing to a population with a nonprobability sample cannot be done strictly with statistical methods. The sample must be assumed to be representative of the population, or a logical, nonstatistical inference must be made. The populations being sampled varied widely from study to study, from multinational to national to local. The population under consideration bears on the ability to generalize the results of a study. Estimates from a sample may provide a point of reference, but are not necessarily normative or representative of a general population or a population of interest.

4.6. Conclusion

The analysis of the NHANES III data that was done here provides a description of the performance of persons in the United States on three computerized neurobehavioral tests. Researchers can use the tables that are provided to compare the performance of populations of interest to results from a national sample. The summary statistics that were calculated can be used as the basis for sample size and power calculations when future studies are designed. In order to avoid confounding, all the demographic variables that are included in the tables should be considered when designing or analyzing data from a study using these three tests. The analysis of the questionnaire data indicates that performance is related to the amount of last night's sleep, energy level, familiarity with computers or video games, alcoholic drinks within the last 3 h, reported effort, the language the test was given in, and test room temperature. These variables should also be considered.

References

[1] J.W. Aird Jr., R.D.G. Webb, J. Hoare, Heat exposure-induced changes in motor outflow component of reaction time, Percept. Motor Skills 56 (1983) 699–706.

[2] L. Altmann, H.-F. Neuhann, U. Krämer, J. Witten, E. Jermann, Neurobehavioral and neurophysiological outcome of chronic low-level tetrachloroethene exposure measured in neighborhoods of dry cleaning shops, Environ. Res. 69 (1995) 83–89.

[3] W.K. Anger, M.G. Cassitto, Y.-X. Liang, R. Amador, J. Hooisma, D.W. Chrislip, D. Mergler, M. Keifer, J. Hörtnagl, L. Fournier, B. Dudek, E. Zsögön, Comparison of performance from three continents on the WHO-recommended Neurobehavioral Core Test Battery, Environ. Res. 62 (1993) 125–147.

[4] W.K. Anger, O.J. Sizemore, S.J. Grossmann, J.A. Glasser, R. Letz, R. Bowler, Human neurobehavioral research methods: Impact of subject variables, Environ. Res. 73 (1997) 18–41.

[5] E.L. Baker, R.G. Feldman, R.A. White, J.P. Harley, C.A. Niles, G.E. Dinse, C.S. Berkey, Occupational lead neurotoxicity — A behavioral electrophysiologic evaluation: I. Study design and year one results, Br. J. Ind. Med. 41 (1984) 352–361.

[6] E.L. Baker, R.E. Letz, E.A. Eisen, L.J. Pothier, D.L. Plantamura, M. Larson, R. Wolford, Neurobehavioral effects of solvents in construction painters, J. Occup. Med. 30 (1988) 116–123.

[7] E.L. Baker, R. Letz, A. Fidler, A computer-administered neurobehavioral evaluation system for occupational and environmental epidemiology: Rationale, methodology and pilot study results, J. Occup. Med. 27 (1985) 206–212.

[8] E.L. Baker, R. Letz, A.T. Fidler, S. Shalat, D. Plantamura, M. Lyndon, A computer-based neurobehavioral evaluation system for occupational and environmental epidemiology: Methodology and validation studies, Neurobehav. Toxicol. Teratol. 7 (1985) 369–377.

[9] K.M. Bang, Health effects of common organic solvents in the workplace, Fam. Community Health 7 (1984) 15–29.

[10] D. Benor, E. Shvartz, Effect of body cooling on vigilance in hot environments, Aerosp. Med. 42 (1971) 727–730.

[11] S.K. Bhattacharya, A. Roy, N.L. Ramanathan, Performance of psychological tests in hot environments, Indian J. Med. Res. 68 (1978) 660–667.

[12] S.K. Bhattacharya, S.R. Tripathi, C.K. Pradhan, S.K. Kashyap, Acute effects of heat on neuropsychological changes and physiological responses under noise condition, Indian J. Exp. Biol. 28 (1990) 849–857.

[13] L. Bolle, H. Herrera, E. Lorétan, M.-A. Boillat, Neurobehavioral test performance among apprentice painters: Baseline data, Am. J. Ind. Med. 29 (1996) 539–546.

[14] R.M. Bowler, C.D. Thaler, D. Law, C.E. Becker, Comparison of the NES and CNS/B neuropsychological screening batteries, Neurotoxicology 11 (1990) 451–464.

[15] D.K. Broadwell, D.J. Darcey, H.K. Hudnell, D.A. Otto, W.K. Boyes, Work-site clinical and neurobehavioral assessment of solvent-exposed microelectronic workers, Am. J. Ind. Med. 27 (1995) 677–698.

[16] G.M. Calvert, C.A. Mueller, J.M. Fajen, D.W. Chrislip, J. Russo, T. Briggle, L.E. Fleming, A.J. Suruda, K. Steenland, Health effects associated with sulfuryl fluoride and methyl bromide exposure among structural fumigation workers, Am. J. Public Health 88 (1998) 1774–1780.

[17] M.G. Cassitto, D. Camerino, H. Hänninen, W.K. Anger, International collaboration to evaluate the WHO Neurobehavioral Core Test Battery, in: B.L. Johnson (Ed.), Advances in Neurobehavioral Toxicology: Applications in Environmental and Occupational Health, Lewis Publishers, Chelsea, MI, 1990, pp. 203–223.

[18] M.G. Cassitto, R. Gilioli, D. Camerino, Experiences with the Milan Automated Neurobehavioral System (MANS) in occupational neurotoxic exposure, Neurotoxicol. Teratol. 11 (1989) 571–574.

[19] M. Colvin, J. Myers, V. Nell, D. Rees, R. Cronje, A cross-sectional survey of neurobehavioral effects of chronic solvent exposure on workers in a paint manufacturing plant, Environ. Res. 63 (1993) 122–132.

[20] C.J. Crespo, S.J. Keteyian, G.W. Heath, C.T. Sempos, Leisure-time physical activity among US adults: Results from the third National Health and Nutrition Examination Survey, Arch. Intern. Med. 156 (1996) 93–98.

[21] C.J. Crespo, C.M. Loria, V.L. Burt, Hypertension and other cardiovascular disease risk factors among Mexican Americans, Cuban Americans, and Puerto Ricans from the Hispanic Health and Nutrition Examination Survey, Public Health Rep. 111 (Suppl. 2) (1996) 7–10.

[22] D. Echeverria, H.V. Aposhian, J.S. Woods, N.J. Heyer, M.M. Aposhian, A.C. Bittner Jr., R.K. Mahurin, M. Cianciola, Neurobehavioral effects from exposure to dental amalgram Hg°: New distinctions between recent exposure and Hg body burden, FASEB J. 12 (1998) 971–980.

[23] D. Echeverria, N.J. Heyer, M.D. Martin, C.A. Naleway, J.S. Woods, A.C. Bittner Jr., Behavioral effects of low-level exposure to Hg° among dentists, Neurotoxicol. Teratol. 17 (1995) 161–168.

[24] A.T. Fidler, E.L. Baker, R.E. Letz, Neurobehavioral effects of occupational exposure to organic solvents among construction painters, Br. J. Ind. Med. 44 (1987) 292–308.

[25] F. Gamberale, A. Iregren, A. Kjellberg, SPES: The Computerized Swedish Performance Evaluation System Background, Critical Issues, Empirical Data, and a Users' Manual, National Institute of Occupational Health, Solna, Sweden, 1989.

[26] B.D. Greenberg, P.A. Moore, R. Letz, E.L. Baker, Computerized assessment of human neurotoxicity: Sensitivity to nitrous oxide exposure, Clin. Pharmacol. Ther. 38 (1985) 656–660.

[27] J.W. Grosch, A.V. Neale, R.Y. Demers, Neurobehavioral and health-related deficits in solvent-exposed painters, Am. J. Ind. Med. 30 (1996) 623–632.

[28] W.R. Holt, E.C. Brainard, Selective hyperthermia and reaction time, Percept. Motor Skills 43 (1976) 375–382.

[29] H.K. Hudnell, D.A. Otto, D.E. House, The influence of vision on computerized neurobehavioral test scores: A proposal for improving test protocols, Neurotoxicol. Teratol. 18 (1996) 391–400.

[30] A.S. Kaufman, J.E. McLean, C.R. Reynolds, Sex, race, residence, region, and education differences on the 11 WAIS-R subtests, J. Clin. Psychol. 44 (1988) 231–248.

[31] E.F. Krieg Jr., D.W. Chrislip, J.M. Russo, A mathematical model of performance on a simple reaction time test, Neurotoxicol. Teratol. 18 (1996) 587–593.

[32] R.J. Kuczmarski, K.M. Flegal, S.M. Campbell, C.L. Johnson, Increasing prevalence of overweight among US adults: The National Health and Nutrition Examination Surveys, 1960 to 1991, J. Am. Med. Assoc. 272 (1994) 205–211.

[33] P. Kumar, S.K. Rastogi, B.N. Gupta, T. Husain, Psychological responses to thermal stress in a glass bangle factory, Occup. Med. 41 (1991) 157–160.

[34] R. Letz, Quantitative neurobehavioral testing in humans for assessing potential effects of occupational exposure, J. Am. Coll. Toxicol. 8 (1989) 303–309.

[35] R. Letz, The neurobehavioral evaluation system: An international effort, in: B.L. Johnson (Ed.), Advances in Neurobehavioral Toxicology: Applications in Environmental and Occupational Health, Lewis Publishers, Chelsea, MI, 1990, pp. 189–201.

[36] R. Letz, Use of computerized test batteries for quantifying neurobehavioral outcomes, Environ. Health Perspect. 90 (1991) 195–198.

[37] R. Letz, Covariates of computerized neurobehavioral test performance in epidemiologic investigations, Environ. Res. 61 (1993) 124–132.

[38] R. Letz, E.L. Baker, Computer-administered neurobehavioral testing in occupational health, Semin. Occup. Med. 1 (1986) 197–202.

[39] R. Letz, R.C. Green, J.L. Woodward, Development of a computer-based battery designed to screen adults for neuropsychological impairment, Neurotoxicol. Teratol. 18 (1996) 365–370.

[40] R. Letz, F.C. Mahoney, D.L. Hershman, S. Woskie, T.J. Smith, Neurobehavioral effects of acute styrene exposure in fiberglass boatbuilders, Neurotoxicol. Teratol. 12 (1990) 665–668.

[41] R. Letz, W.A. Pieper, R.D. Morris, NES test performance in a large US Army veteran sample: Relationships with both demographic factors and traditional neuropsychological measures, Neurotoxicol. Teratol. 18 (1990) 381–390.

[42] Y.-X. Liang, Z.-Q. Chen, R.-K. Sun, Y.-F. Fang, J.-H. Yu, Application of the WHO Neurobehavioral Core Test Battery and other neurobehavioral screening methods, in: B.L. Johnson (Ed.), Advances in Neurobehavioral Toxicology: Applications in Environmental and Occupational Health, Lewis Publishers, Chelsea, MI, 1990, pp. 225–243.

[43] Y.X. Liang, Z.Q. Chen, C. Wang, H.W. Tang, H.G. Yang, The development of neurobehavioral toxicology in China: The 1994 Hänninen Lecture, Environ. Res. 73 (1997) 9–17.

[44] Y.-X. Liang, R.-K. Sun, Y. Sun, Z.-Q. Chen, L.-H. Li, Psychological effects of low exposure to mercury vapor: Application of a computer-administrated neurobehavioral evaluation system, Environ. Res. 60 (1993) 320–327.

[45] B.W. Lovingood, C.S. Blyth, W.H. Peacock, R.B. Lindsay, Effects of d-amphetamine sulfate, caffeine, and high temperature on human performance, Res. Q. 38 (1967) 64–71.

[46] F.C. Mahoney, P.A. Moore, E.L. Baker, R. Letz, Experimental nitrous oxide exposure as a model system for evaluating neurobehavioral tests, Toxicology 49 (1988) 449–457.

[47] N. Maizlish, M. Schenker, C. Weisskopf, J. Seiber, S. Samuels, A behavioral evaluation of pest control workers with short-term, low-level exposure to the organophosphate diazinon, Am. J. Ind. Med. 12 (1987) 153–172.

[48] H. Muijser, R.B.M. Geuskens, J. Hooisma, H.H. Emmen, B.M. Kulig, Behavioral effects of exposure to organic solvents in carpet layers, Neurotoxicol. Teratol. 18 (1996) 455–462.

[49] National Center for Health Statistics, NHANES III Reference Manuals and Reports (CD-ROM), National Center for Health Statistics, Hyattsville, MD, 1996.

[50] National Center for Health Statistics, National Health and Nutrition Examination Survey III, 1988–1994 (CD-ROM Series 11, No. 1A), National Center for Health Statistics, Hyattsville, MD, 1998.

[51] National Center for Health Statistics, National Health and Nutrition Examination Survey III, 1988–1994 (CD-ROM Series 11, No. 2A), National Center for Health Statistics, Hyattsville, MD, 1998.

[52] V. Nell, J. Myers, M. Colvin, D. Rees, Neuropsychological assessment of organic solvent effects in South Africa: Test selection, adaptation, scoring, and validation issues, Environ. Res. 63 (1993) 301–318.

[53] D. Otto, L. Molhave, G. Rose, H.K. Hudnell, D. House, Neurobehavioral and sensory irritant effects of controlled exposure to a complex mixture of volatile organic compounds, Neurotoxicol. Teratol. 12 (1990) 1–4.

[54] D.A. Otto, I. Skalik, D.E. House, H.K. Hudnell, Neurobehavioral Evaluation System (NES): Comparative performance of 2nd-, 4th-, and 8th-grade Czech Children, Neurotoxicol. Teratol. 18 (1996) 421–428.

[55] J.L. Pirkle, D.J. Brody, E.W. Gunter, R.A. Kramer, D.C. Paschal, K.M. Flegal, T.D. Matte, The decline in blood lead levels in the United States: The National Health and Nutrition Examination Surveys (NHANES), J. Am. Med. Assoc. 272 (1994) 284–291.

[56] S. Razmjou, A. Kjellberg, Sustained attention and serial responding in heat: Mental effort in the control of performance, Aviat. Space Environ. Med. 63 (1992) 594–601.

[57] R.E. Reilly, J.F. Parker, Effect of Heat Stress and Prolonged Activity on Perceptual Motor Performance (NASA-CR-1153), Biotechnology, Arlington, VA, 1967.

[58] W.H. Rogers, E.M. Noddin, Manual performance in the cold with gloves and bare hands, Percept. Motor Skills 59 (1984) 3–13.

[59] C.M. Ryan, L.A. Morrow, E.J. Bromet, D.K. Parkinson, Assessment of neuropsychological dysfunction in the workplace: Normative data from the Pittsburgh Occupational Exposures Test Battery, J. Clin. Exp. Neuropsychol. 9 (1987) 665–679.

[60] B.S. Schwartz, K.I. Bolla, W. Stewart, D.P. Ford, J. Agnew, H. Frumkin, Decrements in neurobehavioral performance associated with mixed exposure to organic and inorganic lead, Am. J. Epidemiol. 137 (1993) 1006–1021.

[61] C.T. Sempos, J.I. Cleeman, M.D. Carroll, C.L. Johnson, P.S. Bachorik, D.J. Gordon, V.L. Burt, R.R. Briefel, C.D. Brown, K. Lippel, Preva-

lence of high blood cholesterol among US adults: An update based on guidelines from the second report of the National Cholesterol Education Program Adult Treatment Panel, J. Am. Med. Assoc. 269 (1993) 3009–3014.
[62] E. Shvartz, A. Meroz, A. Mechtinger, H. Birnfeld, Simple reaction time during exercise, heat exposure, and heat acclimation, Aviat. Space Environ. Med. 47 (1976) 1168–1170.
[63] K. Steenland, B. Jenkins, R.G. Ames, M. O'Malley, D. Chrislip, J. Russo, Chronic neurological sequelae to organophosphate pesticide poisoning, Am. J. Public Health 84 (1994) 731–736.
[64] S.-Y. Tsai, J.-D. Chen, Neurobehavioral effects of occupational exposure to low-level styrene, Neurotoxicol. Teratol. 18 (1996) 463–469.
[65] J. Walkowiak, L. Altmann, U. Krämer, K. Sveinsson, M. Turfeld, M. Weishoff-Houben, G. Winneke, Cognitive and sensorimotor functions in 6-year-old children in relation to lead and mercury levels: Adjustment for intelligence and contrast sensitivity in computerized testing, Neurotoxicol. Teratol. 20 (1998) 511–521.
[66] D. Wechsler, WAIS-R Manual Wechsler Adult Intelligence Scale — Revised, Psychological, New York, 1981.
[67] R.F. White, R. Diamond, M. Krengel, K. Lindem, R.G. Feldman, R. Letz, E. Eisen, D. Wegman, Validation of the NES2 in patients with neurological disorders, Neurotoxicol. Teratol. 18 (1996) 441–448.
[68] A.M. Williamson, The development of a neurobehavioral test battery for use in hazard evaluations in occupational settings, Neurotoxicol. Teratol. 12 (1990) 509–514.
[69] H.-G. Yang, Y.-X. Liang, H.-W. Tang, Z.-Q. Chen, P.-G. Li, Neuropsychological effects of lead exposure, Chin. J. Prev. Med. 28 (1994) 9–12.