Assessment
2020, Vol. 27(2) 373–383

DOI: 10.1177/1073191117727346
journals.sagepub.com/home/asm


Evaluation of Normative Data of a Widely Used Computerized Neuropsychological Battery: Applicability and Effects of Sociodemographic Variables in a Dutch Sample

Sophie J. M. Rijnen[1,2], Ikram Meskal[2], Wilco H. M. Emons[2], Carlijn A. M. Campman[2], Sophie D. van der Linden[1,2], Karin Gehring[1,2], and Margriet M. Sitskoorn[2]

Abstract

Introduction: Central Nervous System Vital Signs (CNS VS) is a computerized neuropsychological battery that is translated into many languages. However, published CNS VS' normative data were established over a decade ago, are solely age-corrected, and collected in an American population only. **Method:** Mean performance of healthy Dutch participants on CNS VS was compared with the original CNS VS norms (N = 1,069), and effects of sociodemographic variables were examined. **Results:** z tests demonstrated no significant differences in performance on four out of seven cognitive domains; however, Dutch participants (N = 158) showed higher scores on processing and psychomotor speed, as well as on cognitive flexibility. Although the original CNS VS norms are solely age-corrected, effects of education and sex on CNS VS performance were also identified in the Dutch sample. **Discussion:** Users should be cautious when interpreting CNS VS performance based on the original American norms, and sociodemographic factors must also be considered.

Keywords

CNS Vital Signs, computerized neuropsychological testing, healthy participants, normative data, sociodemographic variables, neuropsychological assessment

Computerized neuropsychological test (CNT) batteries have become increasingly popular in clinical and research settings over the past years. A major advantage of CNT's is the potential of having computers perform labor-intensive test administration, and accurate as well as less time consuming scoring procedures. The Central Nervous System Vital Signs (CNS VS; Gualtieri & Johnson, 2006) is a battery composed of CNTs that are mostly based on well-established conventional paper-and-pencil tests. CNS VS has been shown to be well suited for use as a brief clinical screening tool for cognitive dysfunction in different patient groups (Collins, Mackenzie, Tasca, Scherling, & Smith, 2014; Gualtieri, Johnson, & Benedict, 2006; Meskal, Gehring, van der Linden, Rutten, & Sitskoorn, 2015).

However, in spite of their widespread use and clinical utility, many CNT's, including CNS VS, are limited in terms of their psychometric development, and stratified norms are often lacking (Arrieux, Cole, & Ahrens, 2017; Bauer et al., 2012). Most of the normative data have been collected and described by Gualtieri and Johnson more than a decade ago based on a sample of 1,069 volunteering American participants ranging in age from 7 to 90 years. Since 2006, the normative database has been expanded to over 1,900 participants (http://www.cnsvs.com), but unfortunately no information on the updated CNS VS normative database has been reported to date. As a result, there is no publically available description of the composition of the American sample regarding background characteristics, nor the basis on which participants were classified as "normal," except that they had "no past or present neurological or psychiatric disorder, head injury, and learning disabilities" (Gualtieri & Johnson, 2006, p. 625). Hence, the representativeness of the norms for the American

[1]Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands
[2]Tilburg University, Tilburg, The Netherlands

Corresponding Author:
Sophie J. M. Rijnen, Department of Cognitive Neuropsychology, Tilburg University, Simon Building, Room S201b, P. O. Box 90153, Warandelaan 2, Tilburg 5000 LE, The Netherlands.
Email: s.j.m.rijnen@uvt.nl

population cannot be evaluated and is uncertain. Moreover, although the CNS VS has been translated into over 50 languages, only normative data for the American version has been published. However, the performance on translated versions of the CNS VS could be affected by cultural influences rendering the norms for the American sample inapplicable to individuals in other countries. To the best of our knowledge, the applicability of the original norms to non-American samples has never been studied. In addition, the original CNS VS norms may be outdated, since norms were based on data that were collected over a decade ago. Ageing of norms is an important treat to the usefulness of normative data (e.g., Evers, Sijtsma, Lucassen, & Meijer, 2010).

Another limitation of the original CNS VS' normative data concerns the absence of adjustments for effects of education and sex, as normalized scores are solely age-corrected. All three sociodemographic variables (i.e., age, education, and to a lesser extent sex) have extensively been found to correlate with performance on various neuropsychological tests (Heaton, Grant, & Matthews, 1986; Seidenberg et al., 1984), including performance on computerized tests (Gualtieri & Hervey, 2015; Iverson, Brooks, & Rennison, 2014; Swagerman et al., 2016). The absence of corrections for these variables when interpreting performance on neuropsychological tests hinders proper interpretation and comparison in terms of cognitive functioning.

In the current study, we evaluated the performance of a sample of healthy Dutch participants on the CNS VS against the original American normative data. In addition, we evaluated the impact of the sociodemographic variables age, education, and sex on performance using a regression-based procedure. By using this approach, individual normed scores can be derived. Formulae for obtaining sociodemographically adjusted normed scores based on normative data from the Dutch population are presented as well.

Method

Participants and Procedure

A total of 158 Dutch participants, recruited by convenience sampling from the broad network of the research group, volunteered to participate in the study. Participants were considered healthy if (a) there was no past or present psychiatric or neurologic disorder; (b) they had no other major medical illnesses in the past year prior to participation (e.g., cancer, myocard infarct); (c) they were free of use of any centrally acting psychotropic medication; and (d) did not have a history of or current alcohol or drug abuse. The computerized neuropsychological tests were, depending on participants' preference, administered individually at Tilburg University (Tilburg, The Netherlands), Elisabeth-TweeSteden Hospital (Tilburg, The Netherlands), or at participants' homes. Well-trained test technicians ensured appropriate conditions and remained present during the entire assessment. Participants provided written informed consent and filled out a questionnaire on health status.

The study was approved by the Medical Ethics Committee Brabant, The Netherlands (File number: NL41351.008.12).

Table 1. Description of Educational Levels.

Level	Verhage categories[a]
Low	1. Less than 6 years of primary education
	2. Finished primary education
	3. Primary education and less than 2 years of low-level secondary education
	4. Finished low-level secondary education
Middle	5. Finished average-level secondary education
High	6. Finished high level secondary education
	7. University degree

[a]Adapted from Verhage (1964).

Measures and Normative Data

Sociodemographic Characteristics. Number of years and completed level of education were self-reported by participants. Grade retention did not count as an extra year, neither did supplementary vocational courses that were attended after graduation. Actual number of years of education was verified (i.e., recalculated by the test technician together with the participant) during the assessment. To classify the level of education, the Dutch Verhage scale was used (Verhage, 1964). Its seven categories were merged into three ordinal categories: low educational level (Verhage 1 until 4), middle educational level (Verhage 5), and high educational level (Verhage 6 and 7; Table 1). Participants also rated their frequency of computer use on a 3-point scale with categories *never*, *some*, or *frequent*.

Central Nervous System Vital Signs. Cognitive functioning was assessed using the Dutch translation of the CNT battery CNS VS. It comprises seven neuropsychological tests, yielding measures of performance in 11 cognitive domains. Since some domains scores generated by CNS VS are very similar (i.e., mainly calculated based on components of the same tests), we chose to consider only 7 cognitive domains (Table 2). Time needed to complete the total battery is approximately 30 to 40 minutes. Scoring is automated and scores are presented in raw and normed scores, as well as percentile ranks, generating a summary report for clinical interpretation or statistical analysis. Raw scores include the number of correct or incorrect responses, reflecting accuracy, and mean reaction times (in milliseconds) on individual tests and domains, reflecting speed. Normed scores are automatically generated by the CNS VS and represent the performance of

Table 2. Supplementary Material on Central Nervous System Vital Signs (CNS VS).

Cognitive domain	CNS VS test	Domain score calculations ("Formulas for Calculating Domain Scores," n.d.)	Description
Verbal memory	Verbal memory test (VBM)	VBM direct correct hits + VBM direct correct passes + VBM delayed correct hits + VBM delayed correct passes	Learning a list of 15 words, with a direct recognition, and after six more tests a delayed recognition trial
Visual memory	Visual memory test (VIM)	VIM direct correct hits + VIM direct correct passes + VIM delayed correct hits + VIM delayed correct passes	Learning a list of 15 geometric figures, with a direct recognition, and after six more tests a delayed recognition trial
Processing speed	Symbol digit coding (SDC)	SDC correct responses—SDC errors	Number 1 to 9 correspond to different symbols. As many correct numbers as possible have to be filled out underneath the presented symbols in 90 seconds
Psychomotor speed	Finger-tapping test (FTT); SDC	FTT taps right hand + FTT taps left hand + SDC correct responses	Pressing the space bar with the index finger as many times in 10 seconds, above mentioned
Reaction time	Stroop test (ST)	(ST Part II reaction time on correct responses + ST Part III reaction time on correct responses)/2	In Part I, pressing the space bar as soon as the word RED, YELLOW, BLUE, and GREEN appear—In Part II, pressing the space bar as the color of the word matches what the word says—In Part III, pressing the space bar as the color of the word does not match what the word says
Complex attention	Continuous performance test (CPT); Shifting attention test (SAT); ST	Stroop commission errors + SAT errors + CPT commission errors + CPT omission errors	Responding to a target stimulus "B" but no any other letter. Shifting from one instruction to another quickly and accurately (matching geometric objects either by shape or color); Above mentioned
Cognitive flexibility	SAT; ST	SAT correct—SAT errors—ST commission errors	Above mentioned

an individual relative to the American normative sample controlled for age. In the population, CNS VS normed scores are assumed to have a mean of 100 and a standard deviation of 15; higher scores always indicate better performance (Gualtieri & Johnson, 2006). The percentile rank of these scores refer to the proportion of scores in the normative sample that are equal to or lower than the score at hand. All testing was done using CNS VSX' local software app, on the same type of laptop computers running Windows 7 Professional on 64-bit operating systems. Background programs were shut down at time of all assessments and laptops were disconnected from (wireless) internet resources.

There is not a large body of literature regarding the reliability and validity of CNS VS. In the original reliability and validity paper, Gualtieri and Johnson (2006) describe CNS VS' psychometric characteristics to be very similar to the characteristics of the conventional neuropsychological tests on which the battery is based. However, correlational studies suggest at best moderate correlations between CNS VS and traditional neuropsychological tests, and in addition, no consistent clear patterns of convergent or discriminant validity have been determined (Gualtieri & Hervey, 2015; Gualtieri & Johnson, 2006, 2008; Lanting, Iverson, & Lange, 2012a, 2012b). As no two presentations of CNS VS are similar due to the random presentation of stimuli, the battery is assumed to be suitable for serial administration without inducing practice effects.

CNS VS American Normative Database. As stated before, CNS VS' normative database has been expanded to over 1,900 participants nowadays (http://www.cnsvs.com). However, we rely on information regarding the CNS VS' normative sample described by Gualtieri and Johnson (2006) since detailed information (e.g., sociodemographic characteristics) about the enlarged normative sample is not available.

One thousand sixty-nine normal participants were included in the normative database of CNS VS. Background characteristics (i.e., sex, ethnicity, handedness, and computer familiarity) and normative data are represented for 10 age groups: less than 10 years old, 10 to 14 years, 15 to 19 years, in deciles to 79 years, and finally, 80 years and older, with group sizes ranging from 25 to 212 participants (Gualtieri & Johnson, 2006). In most age groups, there is a female predominance, ranging from 43% to 72%. Characteristics are not presented for the sample as a whole — hindering proper comparisons between the total Dutch

and American samples with respect to age and sex. Information about education (e.g., level, number of years) of the American sample is not described by Gualtieri and Johnson (2006), or in the documentation of the CNS VS itself (http://www.cnsvs.com). Neither was such information available from any of CNS VS' analyses regarding the establishment of the battery's normative data.

Statistical Analysis

Mean Domain and Test Performance. To explore whether mean CNS VS performance of the Dutch participants differed from the mean performance of the normative American sample, a series of two-tailed one-sample *z*-tests was performed (test values: $M = 100$, $SD = 15$). CNS VS presents up to 10 different mean raw scores (i.e., for each of the 10 different age-groups of CNS VS' normative sample) for each domain and test. Since adopting the same subgroups in the Dutch sample would dramatically decrease the sample size for these analyses, the automatically generated age-corrected normed scores were used in all comparisons between the American and Dutch samples. In this way, we also account for effects of age in both groups. Effect sizes (ES) for potential differences between the American and the Dutch samples were calculated and expressed as Cohen's *d* using pooled variance.[1] ES between ≤0.20 and 0.49 were defined as small, between 0.50 and 0.79 as medium, and ≥0.80 represented large effects (Cohen, 1988).

Multiple Regression Analyses. To explore the effects of sociodemographic factors on CNS VS performance, a series of multiple linear regression analyses was conducted using raw CNS VS domain scores as the outcome variables and a predetermined list of sociodemographic predictors. Age (in years), education (dummy coded; middle education as reference category), and sex (coded as 0 = men, 1 = women) were predictor variables which were entered as a single block ("enter" method). Assumptions were evaluated as follows: independence of observations was evaluated by Durbin–Watson tests (Durbin & Watson, 1951), and linearity and homoscedasticity were examined using scatter plots of residuals. Potential multicollinearity between predictors was examined by inspecting Pearson's correlation coefficients. By computing Cook's distances, univariate influential cases were identified (Cook & Weisberg, 1982). Normality of residuals was investigated by visual inspection of histograms. Alpha was set at .02 in order to prevent the problem of inflated Type I errors related to multiple comparisons. All statistical analyses were performed with SPSS 22.0.

Normative Regression Formulae. The results of the regression models which regresses performance on age, sex, and educational level also provide the formulae for computing sociodemographically adjusted norms. Clinicians and researchers can use these formulae in future administrations of CNS VS to obtain normed scores for individuals on each cognitive domain, based on their age, educational level, and sex. In particular, all predictors were included in the normative formulae irrespective of the significance of the effect, as follows:

$$Y_{\text{p domain}} = \alpha + b_1\text{Age} + b_2\text{D}_{\text{low education}} + b_2\text{D}_{\text{high education}} + b_3\text{Sex}$$

In this formula, $Y_{\text{p domain}}$ is the predicted raw domain score, α is the intercept, and b_1 trough b_3 are the regression coefficients. Notice that educational level is a categorical variable with three categories and therefore modeled by means of two dummy variables, one for low education and one for high education (i.e., middle education as reference category). Sex is also a dummy variable, with men as the reference category (i.e., for men: sex = 0 and for women: sex = 1). Application of these regression formulae is demonstrated in Box 1.

Results

Sociodemographic Characteristics

Table 3 shows participants' sociodemographic characteristics. Mean age was 45.9 ($SD = 14.4$) years, ranging from 20.0 to 80.0. There was a female predominance (57%) in the Dutch sample, which appears comparable to the American normative database of CNS VS. The participants completed 16.9 years of education on average. Almost all participants (97%) indicated to use the computer frequently. Men and women did not differ in terms of mean age, $t(156) = 0.48$, $p = .162$, and educational level, $\chi^2(2) = 1.20$, $p = .550$, neither did men and women differ in frequency of computer use, $\chi^2(2) = 1.42$, $p = .491$. Likewise, no significant differences between groups based on the three educational levels were found concerning age, $F(2, 155) = 1.04$, $p = .355$, and frequency of computer use, $\chi^2(4) = 8.79$, $p = .067$.

Mean Domain and Test Performance

Table 4 shows mean differences for the Dutch sample as compared with the American-based normed scores ($M = 100$, $SD = 15$). Significant mean differences were found for the domains of processing speed (mean difference = 4.52, $SD = 14.48$; $z = 3.77$, $p < .001$), psychomotor speed (mean difference = 7.17, $SD = 12.87$; $z = 5.97$, $p <. 001$), and cognitive flexibility (mean difference = 2.91, $SD = 12.94$; $z = 2.39$, $p = .017$), where the Dutch sample demonstrated higher scores than the American normative sample. ES were small (Cohen's *d* respectively 0.19 and 0.30 for cognitive flexibility and processing speed), except for

Box 1. Application of Sociodemographically Adjusted Normative Formulae and a Real-Life Example.

1. Complement the formula: $Y_{p\ domain} = \alpha + b_{1\ age} + b_{2\ low\ education} + b_{2\ high\ education} + b_{3\ sex}$,[a] with the assessed individual's age, education, and sex: this will result in a *predicted raw score* (Y_p) for each cognitive domain.	Consider a 68-year-old male patient who completed a high educational level, and obtained a raw score of 27 on processing speed. His predicted raw score for processing speed is $Y_{p\ processing\ speed} = 77.38 + (-0.52 * age) + (-3.16 * education_{low} + 3.98 * education_{high} + 2.42 * sex_{woman})$, with age = 68, $education_{low}$ = 0, $education_{high}$ = 1, sex_{woman} = 0, resulting in $Y_{p\ processing\ speed}$ = 46.
2. Subtract the predicted raw score from the individual's obtained (Y_o) raw score, now a *difference score* is generated: $Y_o - Y_p$.	The predicted raw score = 46, subtracted from the obtained raw score (27) results in a difference score of −19.
3. The individual's *z* score is computed as follows: *z* score = $Y_o - Y_p/SD_{residual}$, where $SD_{residual}$ is the *SD* of the sample's residual, reflecting the accuracy of predictions made by the regression line.	Dividing the difference score by the $SD_{residual}$ of processing speed = 8.88 (see Table 6), results in *z* = −19/8.88 = −2.14.
4. The *z* score can be interpreted via a *z* distribution. As higher raw scores on reaction time and complex attention indicate worse performance, *z* scores for these domains have to be multiplied by −1 to facilitate consistent interpretation of *z* over all cognitive domains (i.e., positive *z* scores indicate a higher obtained raw score relative to others of similar age, education, and sex, and vice versa for negative *z* scores).	With a *z* score of −2.14, performance on processing speed is more than 2 *SD* lower than expected given the patients' age, education, and sex, which indicates *(serious) impairment*. The obtained raw score of this patient is represented by a CNS VS (age corrected) normed score of 78 (labeled by CNS VS as "below the expected level"), corresponding to a *z* score of: 78−100/15 = −1.47 (as compared to −2.14).

Note. CNS VS = Central Nervous System Vital Signs.
[a]Age in years, sex: 0 = man, 1 = woman; education: low ($education_{low}$ = 1, $education_{high}$ = 0), middle ($education_{low}$ = 0, $education_{high}$ = 0), and high ($education_{low}$ = 0, $education_{high}$ = 1).

Table 3. Sociodemographic Characteristics of the Dutch Sample (*N* = 158) and the American Sample (*N* = 1,069).

	Dutch sample	American sample[a]
Age, years, *M* ± *SD*	45.94 ± 14.43	Unknown[a]
Range	20-80	7-90
Sex, *n* (%)		
Women	90 (57.0)	654 (61.2)
Men	68 (43.0)	415 (38.8)
Education		
Years, *M* ± *SD*	16.88 ± 3.29	Unknown[a]
Level, *n* (%)		
Low	19 (12.0)	Unknown[a]
Middle	57 (36.1)	Unknown[a]
High	82 (51.9)	Unknown[a]
Computer use, *n* (%)		
Never	1 (0.6)	288 (26.9)
Some	4 (2.5)	52 (4.9)
Frequent	153 (96.8)	729 (68.2)

[a]Characteristics of the American sample were not presented for the sample as a whole (see Gualtieri and Johnson [2006] for demographic characteristics across different age groups).

psychomotor speed with a difference of near-medium size (Cohen's $d = 0.49$).

At the level of normed individual test scores (e.g., representing reaction time, number of correct answers), the Dutch sample demonstrated significantly higher scores on 5 out of 17 measures compared with the American normative sample (see Table 4). The number of correct rejections in the delayed recognition Visual Memory Task was significantly higher in the Dutch sample, and Dutch participants performed significantly more taps on the Finger Tapping Test with both the right and the left hand. In addition, the numbers of correct responses on the Symbol Digit Coding task and Shifting Attention Task were higher in the Dutch compared with the original American normative group. A near-medium sized difference was found for the right hand Finger Tapping Test (Cohen's $d = 0.46$), for the other tests, ES were small (Cohen's *d* ranging from 0.20 to 0.36).

Multiple Regression Analyses

None of the assumptions regarding the regression analyses were violated. There was independence of residuals, with Durbin–Watson statistics ranging from 1.72 to 2.22. Scatter plots demonstrated linear relationships between the dependent and independent variables, and homoscedasticity. No problems with collinearity were identified, with correlations *r* between −0.01 and 0.38. No influential cases were identified (all Cook's distances >1), and histograms demonstrated normally distributed standardized residuals for each cognitive domain.

Table 5 shows the results of the regression analyses. Overall, significant effects of age were found on performance in four out of seven raw cognitive domains scores (i.e., for processing speed, psychomotor speed, reaction time, and cognitive flexibility). Higher age was consistently

Table 4. Mean CNS VS Normed Scores of Dutch Participants (N = 158) Compared With the American Normative Data (M = 100; SD = 15).

	M (SD)[a]	Mean difference	z test	P	Effect size d[b]
Domain					
Verbal memory	98.66 (14.99)	−1.34	−1.11	.268	−0.09
Visual memory	101.81 (12.98)	1.81	1.50	.133	0.12
Processing speed	104.52 (14.48)	4.52	3.77	<.001*	0.30
Psychomotor speed	107.17 (12.87)	7.17	5.97	<.001*	0.49
Reaction time	101.41 (11.13)	1.41	1.17	.242	0.09
Complex attention	101.88 (11.66)	1.88	1.54	.124	0.13
Cognitive flexibility	102.91 (12.94)	2.91	2.39	.017*	0.19
Test					
Verbal memory test					
Direct recognition correct hits	99.01 (14.66)	−0.99	−0.79	.425	−0.07
Direct recognition correct rejections	100.94 (12.58)	0.94	0.76	.447	0.06
Delayed recognition correct hits	98.16 (14.86)	−1.84	−1.48	.138	0.12
Delayed recognition correct rejections	98.98 (14.07)	−1.02	−0.89	.370	0.08
Visual memory test					
Direct recognition correct hits	99.50 (13.97)	−0.50	−0.40	.685	0.03
Direct recognition correct rejections	102.53 (13.35)	2.53	2.05	.040	0.17
Delayed recognition correct hits	98.46 (12.06)	−1.54	−1.25	.211	−0.10
Delayed recognition correct rejections	103.86 (11.43)	3.86	3.13	.002*	0.26
Finger-tapping test					
Number of taps right	106.79 (12.66)	6.79	5.52	<.001*	0.46
Number of taps left	104.81 (12.99)	4.81	3.92	<.001*	0.33
Symbol digit coding test					
Number correct	105.37 (14.27)	5.37	4.39	<.001*	0.36
Stroop test					
Reaction time Part I	101.11 (10.01)	1.11	0.91	.364	0.08
Reaction time Part II	100.48 (12.78)	0.48	0.39	.698	0.03
Reaction time Part III	102.34 (10.48)	2.34	1.90	.057	0.16
Shifting attention test					
Number correct	102.97 (14.16)	2.97	2.42	.016*	0.20
Reaction time	100.51 (15.13)	0.51	0.42	.678	0.03
Continuous performance test					
Number correct	101.67 (9.48)	1.67	1.37	.172	0.12

Note. CNS VS = Central Nervous System Vital Signs.
[a]CNS VS normed scores based on the American normative sample have a mean of 100 and a standard deviation of 15; higher scores indicate better performance; positive mean difference indicates better performance for the Dutch sample and vice versa. [b]Cohen's d effect sizes: ≤.20 to .49, small; .50 to .79, medium; ≥.80, large (Cohen, 1988).
*p < .02.

associated with lower scores. Educational level was significantly associated with performance on three out of seven domains: participants with a high educational level (i.e., compared with a middle and low educational level) obtained higher scores on visual memory, processing speed, and cognitive flexibility. Sex was found to be significantly associated with performance on the verbal memory domain, in favor of women, and the psychomotor speed domain, in favor of men. The proportions of explained variances (R^2) by age, education, and sex ranged from 7.2% (for the verbal memory domain) up to 46.2% (for the processing speed domain). Hierarchical regression analyses demonstrated significantly more explained variance for a model including both age and education, compared with a model with solely age, in four out of seven cognitive domains. In two out of seven domains adding the factor sex on top of age and education resulted in significantly more variance explained. Adding education or sex (in addition to age) to the regression model significantly increased the explained variance for the cognitive domains, except for the reaction time domain, where only age contributes significantly (data not shown).

Table 5. Multiple Regression Based on the Dutch Sample (*N* = 158): Association of Age, Education, and Sex With Raw Cognitive Domain Scores of CNS VS.

				95% CI B				
Cognitive domain	Predictor	B	*SE* B	Lower limit	Upper limit	*p*	*F*(*df*)	R^2
Verbal memory						<.001*	2.91(4)	.072
	Age	−0.03	0.03	−0.08	0.03	.320		
	$Education_{low}$	−0.21	1.21	−2.61	2.18	.861		
	$Education_{high}$	1.49	0.79	−0.07	3.05	.062		
	Sex_{woman}	1.77	0.74	0.31	3.23	.018*		
Visual memory						<.001*	4.55(4)*	.108
	Age	−0.06	0.02	−0.10	−0.01	.021		
	$Education_{low}$	−0.92	1.12	−3.13	1.29	.415		
	$Education_{high}$	1.79	0.73	0.35	3.22	.015*		
	Sex_{woman}	0.83	0.68	−0.51	2.18	.222		
Processing speed						<.001*	32.83(4)*	.462
	Age	−0.52	0.05	−0.62	−0.42	<.001*		
	$Education_{low}$	−3.16	2.41	−7.91	1.60	.191		
	$Education_{high}$	3.98	1.56	0.92	7.06	.011*		
	Sex_{woman}	2.42	1.45	−0.45	5.29	.097		
Psychomotor speed						<.001*	24.49(4)*	.392
	Age	−0.87	0.09	−1.05	−0.68	<.001*		
	$Education_{low}$	0.22	4.5	−8.65	9.09	.960		
	$Education_{high}$	5.56	2.91	−0.18	11.31	.058		
	Sex_{woman}	−7.44	2.72	−12.82	−2.07	.007*		
Reaction time[a]						<.001*	6.22(4)*	.142
	Age	1.65	0.39	0.88	2.42	<.001*		
	$Education_{low}$	−11.43	19.09	−49.17	26.30	.550		
	$Education_{high}$	−20.51	12.11	−44.43	3.42	.092		
	Sex_{woman}	−20.71	11.36	−43.16	1.73	.070		
Complex attention[a]						<.001*	4.11(4)*	.100
	Age	0.03	0.02	−0.02	0.08	.189		
	$Education_{low}$	2.60	1.17	0.29	4.90	.027		
	$Education_{high}$	−1.30	0.73	−2.75	0.14	.076		
	Sex_{woman}	0.65	0.69	−0.70	2.01	.343		
Cognitive flexibility						<.001*	12.38(4)*	.249
	Age	−0.28	0.06	−0.40	−0.17	<.001*		
	$Education_{low}$	−6.16	2.82	−11.76	−0.57	.031		
	$Education_{high}$	5.05	1.81	1.48	8.62	.006*		
	Sex_{woman}	−1.63	1.69	−4.98	1.72	.337		

Note. CNS VS = Central Nervous System Vital Signs; *df* = degrees of freedom; *SE* B = standard error B; 95% CI B = 95% confidence interval B. Coding of predictors: age in years; low level of education: $eduction_{low}$ = 1, $education_{high}$ = 0; middle level of education: $education_{low}$ = 0, $education_{high}$ = 0; high level of education: $education_{low}$ = 0, $education_{high}$ = 1; sex: man = 0, woman = 1.
[a]Higher scores indicate lower performance.
*p < .02.

Normative Regression Formulae

Table 6 shows the regression formulae that can be used to calculate normed predicted scores (i.e., corrected for effects of age, education, and sex) on cognitive domains of CNS VS for the Dutch population. An example of the application of the sociodemographically adjusted normative formulae is shown in Box 1.

Discussion

We examined the performance of a group of healthy Dutch participants who underwent neuropsychological examination with the computerized neuropsychological battery CNS VS. The purpose of this study was threefold: (a) to examine the applicability of the American CNS VS norms for the Dutch population; (b) to examine the effects of age,

Table 6. Regression Formulae Based on the Dutch Sample (N = 158).

Cognitive domain	Regression equation[a]	$SD_{residual}$
Verbal memory	50.93 + (−0.03 * age) + (−0.21 * $education_{low}$ + 1.49 * $education_{high}$) + (**1.77 * sex_{woman}**)	4.47
Visual memory	47.33 + (−0.06 * age) + (−0.92 * $education_{low}$ + **1.79 * $education_{high}$**) + (0.83 * sex_{woman})	4.12
Processing speed	77.38 + (**−0.52 * age**) + (−3.16 * $education_{low}$ + **3.98 * $education_{high}$**) + (2.42 * sex_{woman})	8.88
Psychomotor speed	219.00 + (**−0.87 * age**) + (0.22 * $education_{low}$ + 5.56 * $education_{high}$) + (**−7.44 * sex_{woman}**)	16.58
Reaction time[b]	590.03 + (**1.65 * age**) + (−11.43 * $education_{low}$ + −20.51 * $education_{high}$) + (−20.71 * sex_{woman})	69.03
Complex attention[b]	5.07 + (0.03 * age) + (2.60 * $education_{low}$ + −1.30 * $education_{high}$) + (0.65 * sex_{woman})	4.13
Cognitive flexibility	58.51 + (**−0.28 * age**) + (−6.16 * $education_{low}$ + **5.05 * $education_{high}$**) + (−1.63 * sex_{woman})	10.21

Note. Age in years, sex: 0 = man and 1 = woman; education: low ($education_{low}$ = 1, $education_{high}$ = 0), middle ($education_{high}$ = 0, $education_{low}$ = 0), and high ($education_{low}$ = 0, $education_{high}$ = 1). $SD_{residual}$ = standard deviation of the sample's residual. $p < .02$ in bold.
[a]$Y_{p\ domain} = - + b_1 Age + b_2 D_{low\ education} + b_2 D_{high\ education} + b_3 Sex$. [b]Higher scores on Reaction time and Complex attention indicate worse performance z scores for these domains have to be multiplied by −1 to facilitate consistent interpretation over all cognitive domains.

education, and sex on CNS VS performance of the Dutch sample; and (c) to provide sociodemographically adjusted normative formulae for the Dutch population.

At the level of individual CNS VS tests, scores in the Dutch sample were significantly higher on 5 out of 17 measures. Consequently, differences in mean performance for three out of seven cognitive domains were found between the Dutch sample and the American normative sample; in the two domains covering different types of speed, namely processing and psychomotor speed, and in cognitive flexibility.

It should be noted that computer skills — including keyboard work and on-screen visual scanning — have improved tremendously over the past decade, which may result in improvements in overall performance on computerized neuropsychological speed tests. Indeed, an earlier study on computer familiarity and CNS VS performance demonstrated significantly better (i.e., faster) performance in people who are very familiar with computers, opposed to people who reported only "some" familiarity with computers (Iverson, Brooks, Ashton, Johnson, & Gualtieri, 2009). As can be expected from the more frequent use of computers nowadays, our sample comprised too few participants with only some or none computer familiarity to look into these effects. The beneficial effects of computer familiarity may (partly) explain the differences between the American 2006 group and the Dutch 2016 group.

In addition, a possible Flynn effect should be considered given the headspring of the normative data presented by CNS VS. The Flynn effect refers to a substantial rise of the population's performance on tests of intelligence in developed countries, typically about 3 to 5 points (i.e., on a IQ scale with a mean of 100 and standard deviation of 15 points) per decade. Explanations for the Flynn effect include genetic, environmental, methodological, and measurement factors (Flynn, 1984; Trahan, Stuebing, Hiscock, & Fletcher, 2014). It has been found that the impact of the Flynn effect extends beyond the measurement of IQ and has, for example, been demonstrated on measures of memory (Baxendale, 2010; Rönnlund & Nilsson, 2009), processing speed, and cognitive flexibility (Dickinson & Hiscock, 2011), with gains comparable to the size of the Flynn effect on measures of IQ. The scale of normed scores of CNS VS tests and domains is similar to that of IQ points, and the original normative data presented by Gualtieri and Johnson (2006) have been established over a decade ago. Therefore, mean normed cognitive domain scores can be expected to be about 3 to 5 points higher in the current 2015/2016 sample than the original normative data—which corresponds to the increased scores found in the present study.

Since the total variance explained by the sociodemographic variables added up to almost 50% in the present study (i.e., in particular for the processing speed and psychomotor speed domain), the influence of age, sex, and education should be taken into consideration when interpreting performance on the CNS VS. CNS VS incorporated corrections for age in their normative evaluation, but did not correct for effects of education and sex. Consistent with the literature, higher age was associated with lower performance (Verhaegen & Salthouse, 1997). Educational level was found to be positively associated with performance on visual memory, processing speed, and cognitive flexibility. Highly educated participants are likely to be somewhat overrepresented in our Dutch sample relative to the general Dutch population (CBS Statistics Netherlands, http://statline.cbs.nl/Statweb/). Although the higher performance of the sample might also be

explained by this factor, we have no information on education in the original American normative sample, as these data are not disclosed by the authors. We may assume that this sample also included a relatively high proportion of highly educated participants, as these are typically (more) interested in study participation (Wacholder, Silverman, McLaughin, & Mandel, 1992). As would be expected, sex did not play a large role, except for the verbal memory domain favoring women, and the psychomotor speed domain favoring men. These findings are consistent with literature on sex differences in performance on other (computerized) tests (Gur et al., 2001; Iverson et al., 2014; Lezak, Howieson, Loring, Hannay, & Fischer, 2004; Silverstein et al., 2007), and reported by Gualtieri (n.d.) who examined sex differences in a subset of participants who completed the CNS VS battery during its standardization study.

Based on the collected data, we established regression-based normative formulae to adjust for the effect of sociodemographic variables on CNS VS performance. In future evaluations of performance in our (Dutch) patient studies, these normative data will replace the American norms.

Some critical remarks are in order with respect to the current study. Presented results are based on performance in Dutch healthy participants recruited on availability (i.e., convenience sampling). A disadvantage of this method includes the risk that the sample might not represent the Dutch population as a whole. As stated above, a relatively small number of low-educated participants (i.e., 12% compared with approximately 35% in the general Dutch population (CBS Statistics Netherlands, http://statline.cbs.nl/Statweb/) was included in the present study. The regression-based method requires smaller samples since continuous covariates do not have to be categorized (e.g., stratifying the sample into groups of different age, sex, and educational levels). Instead, it makes optimum use of the entire sample to estimate the normative statistics and the regression model (Evers, Lucassen, Meijer, & Sijtsma, 2009; Oosterhuis, van der Ark, & Sijtsma, 2016). However, one should always be careful when using these data for interpreting individual test performance of people who are in the extreme ends of age, or education (very low levels, or by contrast, very high levels of education). In addition, data were collected using a Dutch translation of the CNS VS battery. Since the equivalence with the English version of the test has never been confirmed, we cannot rule out that differences in difficulty due to translation of instructions and items also have a share in the observed differences (Bender, García, & Barr, 2010). Although our results may not be generalizable to other countries or to populations who speak other languages, they demonstrate that CNS VS users from other populations than the American should use and interpret the original norms with caution. Moreover, we recommend on considering adjusting for sociodemographic factors when interpreting CNS VS performance in American populations.

Also, changes in technology (i.e., computer hardware/software) since the collection of the initial American norms may have affected important parameters including timing accuracy. Although technical aspects and settings of the devices used in the present study were the same for all assessments, no information is available concerning devices that were used when collecting the American normative data. Differences therein might explain a small portion of the group differences in our study, but this is unlikely considering the generally rather small timing inaccuracies and the significant differences that we demonstrated for cognitive domains (Cernich, Brennana, Barker, & Bleiberg, 2007; Plant & Turner, 2009). Yet, although the timing precision of CNS VS on different test systems should be explored in more detail, CNS VS provides explicit recommendations concerning system requirements for installation, and states that their applications are designed to be working equally well over types of devices and types of applications (‘CNS Vital Signs Optimal Use Installation Guide’, www.cnsvs.com). However, evidence of this statement is not available.

Future studies should consider the psychometric robustness of CNS VS across cultures and (other) non-American languages. Furthermore, various clinical and research settings require repeated neuropsychological assessment, for example, for the evaluation of effects of intervention on cognitive functioning. This emphasizes the need for inspection of CNS VS regarding repeated assessment, addressing practice effects (i.e., improvements in performance due to familiarity with the test, its items, and test procedures opposed to true cognitive improvement). Currently, we are performing follow-up assessments in the same Dutch sample with the aim of establishing change indices correcting for potential practice effects and measurement errors to determine “true” (i.e., reliable) clinically meaningful cognitive change when administering CNS VS repeatedly over time.

The present study examined the applicability of the original American normative data of CNS VS to a non-American population: our results call the usefulness of the 2006 norms of the CNS VS in other populations than the American into question. Furthermore, we identified effects of education and sex, in addition to known effects of age, on CNS VS performance. These findings highlight the need for more up-to-date population-based norms for CNS VS performance. Sociodemographic factors should be considered when interpreting performance on this measure, for example, by applying sociodemographically adjusted normative formulae, as we have presented here.

Acknowledgments

The authors thank the research assistants for the recruitment and neuropsychological assessment, and in addition, Eline Verhaak and Wietske Schimmel for comments on an earlier version of this article.

Authors' Note

A SPSS syntax for converting raw cognitive domain scores into *z* scores is available by contacting the authors.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is funded by ZonMw, a Dutch national organization for Health Research and Development. Project Number 842003007.

Note

1. $M_1 - M_2 \,/\, SD_{\text{pooled}}$, with $SD_{\text{pooled}} = \sqrt{(n_1-1)SD_1^2 + (n_2-1)SD_2^2 \,/\, (n_1-1) + (n_2-1)}$.

References

Arrieux, J. P., Cole, W. R., & Ahrens, A. P. (2017). A review of the validity of computerized neurocognitive assessment tools in mild traumatic brain injury assessment. *Concussion*, *2*. Retrieved from https://www.futuremedicine.com/doi/10.2217/cnc-2016-0021

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist*, *26*, 177-196. doi:10.1080/13854046.2012.663001

Baxendale, S. (2010). The Flynn effect and memory function. *Journal of Clinical and Experimental Neuropsychology*, *32*, 699-703. doi:10.1080/13803390903493515

Bender, H.A., García, A.M., & Barr, W.B. (2010). An interdisciplinary approach to neuropsychological test construction: Perspectived from translation studies. *Journal ot the International Neuropsychological Society*, *16*, 227-232. doi: 10.1017/S1355617709991378

Cernich, A.N., Brennana, D.M., Barker, L.M., & Bleiberg, J. (2007). Sources of error in computerized neuropsychological assessment. *Archives of Clinical Neuropsychology*, *22S*, S39-S48. doi: 10.1016/j.acn.2006.10.004

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Collins, B., Mackenzie, J., Tasca, G. A., Scherling, C., & Smith, A. (2014). Persistent cognitive changes in breast cancer patients 1 year following completion of chemotherapy. *Journal of the International Neuropsychological Society*, *20*, 370-390. doi:10.1017/S1355617713001215

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York, NY: Chapman & Hall.

Dickinson, D., & Hiscock, H. (2011). The Flynn effect in neuropsychological assessment. *Applied Neuropsychology*, *18*, 136-142. doi:10.1080/09084282.2010.547785

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression: II. *Biometrika*, *38*, 159-177. doi:10.2307/2332325

Evers, A., Lucassen, W., Meijer, R. R., & Sijtsma, K. (2009). *COTAN beoordelingssyteem voor de kwaliteit van tests* [COTAN assessment system for the quality of tests]. Amsterdam, Netherlands: Nederlands Instituut van Psychologen.

Evers, A., Sijtsma, K., Lucassen, W., & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure, and results. *International Journal of Testing*, *10*, 295-317. doi:10.1080/15305058.2010.518325

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*, 29-51. doi:10.1037/0033-2909.95.1.29

Formulas for Calculating Domain Scores. (n.d.). Retrieved from http://www.cnsvs.com/WhitePapers/CNSVS-BriefInterpretationGuide.pdf

Gualtieri, C.T. (n.d.). *Gender differences in cognitive ability from age 40 to 79*. Retrieved from http://www.ncneuropsych.com

Gualtieri, C. T., & Hervey, A. S. (2015). The structure and meaning of a computerized neurocognitive battery. *Frontiers in Psychological and Behavioral Science*, *4*(2), 11-21.

Gualtieri, C. T., & Johnson, L. G. (2006). Reliability and validity of a computerized neurocognitive battery, CNS Vital Signs. *Archives of Clinical Neuropsychology*, *21*, 623-643. doi:10.1016/j.acn.2006.05.007

Gualtieri, C. T., & Johnson, L. G. (2008). A computerized test battery sensitive to mild and severe brain injury. *Medscape Journal of Medicine*, *10*, 90.

Gualtieri, C. T., Johnson, L. G., & Benedict, K. B. (2006). Neurocognition in depression: Patients on and of medication versus healthy comparison subjects. *Journal of Neuropsychiatry and Clinical Neurosciences*, *18*, 217-225. doi:10.1176/jnp.2006.18.2.217

Gur, R.C., Ragland, J.D., Moberg, P.J., Turner, T.H., Bilker, W.B., Kohler, C., . . . Gur, R.E. (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology*, *25*(5), 766-776.

Heaton, R. K., Grant, I., & Matthews, C. G. (1986). Differences in neuropsychological test performance associated with age, education, and sex. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment in neuropsychiatric disorders: Clinical methods and empirical findings* (pp. 100-120). Oxford, England: Oxford University Press.

Iverson, G. L., Brooks, B. L., Ashton, V. L., Johnson, L. G., & Gualtieri, C. T. (2009). Does familiarity with computers affect computerized neuropsychological test performance? *Journal of Clinical and Experimental Neuropsychology*, *31*, 594-604. doi:10.1080/13803390802372125

Iverson, G. L., Brooks, B. L., & Rennison, V. L. A. (2014). Minimal gender differences on the CNS Vital Sings computerized neurocognitive battery. *Applied Neuropsychology: Adult*, *21*, 36-42. doi:10.1080/09084282.2012.721149

Lanting, S. C., Iverson, G. L., & Lange, R. T. (2012a, October). *Comparing patients with mild traumatic brain injury to trauma controls on CNS Vital Signs*. Poster presented at the American Congress of Rehabilitation Medicine Conference, Vancouver, British Columbia, Canada.

Lanting, S. C., Iverson, G. L., & Lange, R. T. (2012b, October). *Concurrent validity of CNS Vital Signs in patients with mild traumatic brain injury*. Poster presented at the American Congress of Rehabilitation Medicine Conference, Vancouver, British Columbia, Canada.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment* (4th ed.). Oxford, England: Oxford University Press.

Meskal, I., Gehring, K., van der Linden, S. D., Rutten, G.-J. M., & Sitskoorn, M. M. (2015). Cognitive improvement in meningioma patients after surgery: Clinical relevance of computerized testing. *Journal of Neuro-Oncology*, *121*, 617-625. doi:10.1007/s11060-014-1679-8

Plant, R.R., & Turner, G. (2009). Millisecond precision psychological research in a world of commodity computers: New hardware, new problems? *Behavior Research Methods*, *41*(3), 598-614. doi: 10.3758/BRM.41.3.598

Oosterhuis, H. E. M., van der Ark., L. A., & Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, *23*, 191-200. doi:10.1177/1073191115580638

Rönnlund, M., & Nilsson, L.-G. (2009). Flynn effects on sub-factors of episodic and semantic memory: Parallel gains over time and the same set of determining factors. *Neuropsychologia*, *47*, 2174-2180. doi:10.1016/j.neuropsychologia.2008.11.007

Seidenberg, M., Gamache, M. P., Beck, N. C., Giordani, B., Berent, S., Sackellares, J. C., . . . Boll, T. J. (1984). Subject variables and performance on the Halstead Neuropsychological Test Battery: A multivariate analysis. *Journal of Consulting and Clinical Psychology*, *52*, 658-662. doi:10.1037/0022-006X.52.4.658

Silverstein, S.M., Berten, S., Olson, P., Paul, R., Williams, L.M., Cooper, N., . . . Gordon, E. (2007). Development and validation of a World-Wide-Web-based neurocognitive assessment battery: WebNeuro. *Behavior Research Methods*, *39*(4), 940-949. doi: 10.3758/BF03192989

Swagerman, S. C., de Geus, E. J. C., Kan, K. J., van Bergen, E., Nieuwboer, H. A., Koenis, M. M. G., . . . Boomsma, D. I. (2016). The computerized neurocognitive battery: Validation, aging effects, and heritability across cognitive domains. *Neuropsychology*, *30*, 53-64. doi:10.1037/neu0000248

Trahan, L., Stuebing, K. K., Hiscock, M. K., & Fletcher, J. M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin*, *140*, 1332-1360. doi:10.1037/a0037173

Verhaegen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, *122*, 231-249. doi:10.1037/0033-2909.122.3.231

Verhage, F. (1964). *Intelligentie en leeftijd onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar* [Intelligence and age: Research study in Dutch individuals aged twelve to seventy-seven]. Assen, Netherlands: Van Gorcum/Prakke & Prakke.

Wacholder, S., Silverman, D. T., McLaughin, J. K., & Mandel, J. S. (1992). Selection of controls in case-controls studies: II: Types of controls. *American Journal of Epidemiology*, *135*, 1029-1041. doi:10.1093/oxfordjournals.aje.a116397