ORIGINAL ARTICLE

Copyright © 2017 The Canadian Journal of Neurological Sciences Inc.

Canadian Normative Data for Minimal Assessment of Cognitive Function in Multiple Sclerosis

Lisa A.S. Walker, David Marino, Jason A. Berard, Anthony Feinstein, Sarah A. Morrow, Denis Cousineau

ABSTRACT: ***Objective:*** The Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS) is a consensus-based collection of neuropsychological tests that evaluate cognitive functioning in individuals with multiple sclerosis (MS). The tests are typically scored using each respective published test manual, leaving the examiner to make interpretations from norms derived from different American populations. Given demographic differences, this may lead to misinterpretation of findings in Canadians. Our goal was to establish both discrete and regression-based normative data for the MACFIMS based on a largely co-normed Canadian population to allow for improved psychometric interpretation. ***Methods:*** MACFIMS data sets were aggregated from across three different Canadian cities (Ottawa, Toronto, and London), yielding a total of 330 healthy control participants from four different studies evaluating cognition in individuals with MS. Given the variety of contributing studies, there was variability in terms of the number of participants completing each measure. ***Results:*** Both age-based discrete normative data and demographically adjusted (sex, age, and education) regression-based formulae were established. The demographic variables varied in their contribution to each MACFIMS test in the regression models, predicting 0 to 18% of the variance. ***Conclusions:*** Provision of these regression-based formulae will allow for more accurate interpretation of Canadian-derived MACFIMS scores by allowing clinicians to correct for all relevant demographic variables simultaneously, leading to improved clinical decision making for individuals with multiple sclerosis.

RÉSUMÉ: **Données normatives canadiennes pour le *Minimal Assessment of Cognitive Function in Multiple Sclerosis*.** ***Objectif***: Le *Minimal Assessment of Cognitive Function in Multiple Sclerosis* (MACFIMS) est un ensemble de tests neuropsychologiques, basé sur un consensus, qui évaluent la fonction cognitive d'individus atteints de sclérose en plaques (SP). Les tests sont habituellement notés selon les directives contenues dans les manuels publiés pour chacun des tests, laissant l'examinateur en faire l'interprétation à partir de normes provenant de différentes populations américaines. Étant donné l'existence de différences démographiques entre différentes populations, cette façon de procéder peut mener à une interprétation erronée des observations chez des sujets canadiens. Le but de notre étude était d'établir des données normatives discrètes ainsi que des données s'appuyant sur une analyse de régression pour le MACFIMS, basées sur une population canadienne de référence, afin d'en améliorer l'interprétation psychométrique. ***Méthodologie***: Des ensembles de données du MACFIMS provenant de trois villes canadiennes (Ottawa, Toronto et London) ont été regroupés, soit au total les données de 330 sujets témoins en bonne santé qui avaient participé à 4 études différentes évaluant la fonction cognitive d'individus atteints de SP. Étant donné la diversité des études, le nombre de participants qui avaient complété chaque mesure pouvait varier. ***Résultats***: Nous avons établi des formules basées sur des données normatives discrètes, selon l'âge, et sur des données ajustées au point de vue démographique (le sexe, l'âge et le niveau de scolarité) à l'analyse de régression. La contribution des variables démographiques à chaque test du MACFIMS variait dans les modèles de régression, prédisant de 0 à 18% de la variance. ***Conclusions***: Des formules basées sur l'analyse de régression permettront une interprétation plus juste des scores canadiens au MACFIMS parce que les cliniciens pourront corriger simultanément les données pour toutes les variables démographiques pertinentes, ce qui entraînera une amélioration des décisions cliniques concernant les patients atteints de MS.

Keywords: Multiple sclerosis, cognition, neuropsychological tests, psychometrics, Canada
doi:10.1017/cjn.2017.199 Can J Neurol Sci. 2017; 44: 547-555

A number of different assessment tools have been proposed over the past few decades to evaluate cognition in multiple sclerosis (MS).[1-4] However, the most commonly accepted comprehensive battery of tests for neuropsychological evaluation in MS is the Minimal Assessment of Cognitive Function in MS (MACFIMS).[5] This battery of tests was chosen via consensus by

From the Ottawa Hospital Research Institute, Ottawa, Ontario, Canada (LASW, JAB); University of Ottawa Brain and Mind Research Institute, Ottawa, Ontario, Canada (LASW); School of Psychology, University of Ottawa, Ottawa, Ontario, Canada (LASW, JAB, DC); Department of Psychology and Institute of Cognitive Science, Carleton University, Ottawa, Ontario, Canada (LASW); Department of Psychology, University of Toledo, Toledo, Ohio, USA (DM); Sunnybrook Health Sciences Centre and University of Toronto, Toronto, Ontario, Canada (AF); London Health Sciences Centre and Western University, London, Ontario, Canada (SAM).

Received January 18, 2017. Final Revisions Submitted March 13, 2017. Date of Acceptance April 1, 2017.

Correspondence to: Lisa A.S. Walker, The Ottawa Hospital, 501 Smyth Road, Suite 7300, Ottawa, Ontario K1H 8L6, Canada. E-mail: lwalker@toh.ca.

https://doi.org/10.1017/cjn.2017.199 Published online by Cambridge University Press

an expert panel of neuropsychologists and psychologists (with clinical and research expertise in MS) from the United States, Canada, United Kingdom, and Australia at a 2001 meeting sponsored by the Consortium of MS Centers (CMSC).[5] Following a review of the pertinent literature, criteria were established with which to select appropriate measures. These criteria included: standardized stimulus materials and administration, normative data, adequate range, reliability, criterion validity, alternate forms and practicality.[5] In addition, the measures chosen had to reflect the domains of cognition typically affected in MS: processing speed/working memory, language, visual perception/spatial processing, learning and memory, and executive function. The following measures were selected: the Paced Auditory Serial Addition Test (Rao version; PASAT);[6,7] the oral Symbol Digit Modalities Test (Rao version; SDMT);[6,8] the Controlled Oral Word Association Test;[9] the Judgment of Line Orientation Test (JLO);[10] the California Verbal Learning Test–II (CVLT–II);[11] the Brief Visuospatial Memory Test–Revised (BVMT–R);[12] and the Delis–Kaplan Executive Function System (D–KEFS) Sorting Test (Sorting).[13] This battery has been well-validated in MS and is used in common practice by clinicians assessing cognition in MS.[14] Components of the MACFIMS battery (i.e., the BVMT–R) are more sensitive to impairment than other established batteries[15] and are predictive of health-related quality of life (i.e., the SDMT).[16] MACFIMS data are strongly related to the neuropsychiatric features of the disease, neuroimaging, and vocational outcomes.[17-19]

When deciding whether or not a patient is cognitively impaired on the MACFIMS, clinicians typically consult normative data for each respective test. These data are provided in all published test manuals, allowing clinicians to compare patients' performance relative to that of demographically similar healthy individuals. Typically, performance is considered impaired if a person scores at, or more than, 1.5 standard deviations below the normative mean. Given that age affects performance on most cognitive tasks,[20] normative data are typically supplied in predetermined, and sometimes arbitrary, age groupings. Performance may also be affected by educational level and gender, and thus some normative data take these variables into account as well.[21,22] Such *discrete* norms are typically used in common practice by clinicians.[23-25]

More recently, there is a movement toward *regression-based* or *continuous* norms.[26] Data are derived from regression-based formulae in which numerous demographic variables can be included (e.g., age, education, sex, race, IQ, and physical health).[27,28] Regression-based norms are not vulnerable to the same limitations plaguing discrete norms such as arbitrary cutoffs for the age/education groupings (which can affect interpretation depending on which age category a person is assigned to),[29] small numbers of participants in individual subgroups, and the lack of correction for all relevant demographic information.[30] In addition, typically smaller sample sizes are needed in regression-based normative data to obtain norms as precise as those provided by discrete methods.[29]

To the best of our knowledge, although several studies have presented control data on the MACFIMS when comparing a full control sample to an MS sample,[14,31,32] these data have not been presented in a manner that would allow their use for normative comparison in a clinical setting. Indeed, the control data are presented as a group, rather than broken down into age or other demographically relevant categories. Thus, there are no clinically relevant discrete norms available for the MACFIMS on a largely co-normed population. In other words, although discrete published normative data are available for all components of the MACFIMS, each particular test was normed on different populations. Although some researchers suggest that co-norming is not required for competent clinical practice,[33] others stress that norms derived from different populations cannot be considered equivalent given differences in demographics between the various normative populations from which individual published tests were derived.[34] Without co-norming, there is room for error in interpretation given that the populations are not likely to be equivalent. Despite this, clinicians are increasingly relying on flexible batteries, so that in practice co-norming is not considered a priority.[33] Nonetheless, given the psychometric advantage of co-norming, having such data available for the MACFIMS would be an asset.

In the area of cognitive assessment in MS, regression-based normative data are used with the National MS Society Consensus Neuropsychological Battery for Pediatric Multiple Sclerosis (NBPMS)[35,36] and have also been derived for the MACFIMS in an American sample.[37] The latter was demonstrated to yield higher rates of impairment than manualised norms for many of the MACFIMS measures when evaluating people with MS. Although members of our group have published regression-based normative data on the SDMT in a Canadian sample[30] (and the Ottawa control data from that study are also included in the current paper), to the best of our knowledge, no such Canadian normative data are available on the full MACFIMS battery. There is a need for such data given that it has been demonstrated that Canadian adults attain higher raw scores than Americans on measures of general intellectual ability as evaluated by the Wechsler Adult Intelligence Scale (versions III and IV),[38] which has been attributed to variations in population composition due to social, economic, and educational differences between the two countries.[39] When assessing a person with an acquired neurological condition, clinicians may conclude that the person is more significantly impaired when using Canadian norms relative to American ones.[39] Others suggest that education level has a stronger relationship with IQ in American samples compared to Canadian ones.[40] Miller et al. (2015)[38] concluded that "choosing to use U.S. norms to interpret the test scores of a Canadian will increase the variance and subsequently increase the [confidence intervals] around the true score, making it more difficult to rely on the test scores for diagnosis" (P. 323–4). This highlights the need for Canadian MACFIMS normative data and is reflective of best practice recommendations.[38] MS researchers from other countries are also recognizing the need for culturally sensitive use of MS-focused testing batteries such as the MACFIMS[31,32] and the Brief International Cognitive Assessment for MS (BICAMS).[41-44]

Our current objective was to establish both discrete and regression-based normative data for the MACFIMS based on a partially co-normed Canadian population.

Methods

The studies that contributed to this database were approved by appropriate institutional review boards, which included the Ottawa Health Science Network Research Ethics Board, the Sunnybrook Health Sciences Centre Research Ethics Board, and the Health Sciences Research Ethics Board (London).

https://doi.org/10.1017/cjn.2017.199 Published online by Cambridge University Press

Participants

Research participants were healthy controls from three medium to large Ontario cities (Ottawa, London, and Toronto). Participants were recruited from the community, through word of mouth, via posted advertisements, and using website advertisements. All participating subjects provided full informed consent. The data were collected in the course of four different studies (each with different original aims) evaluating various aspects of cognitive functioning in persons with multiple sclerosis (see Supplementary Table 1). Overall, 174 participants were obtained from studies conducted by Dr. Walker in Ottawa. The Toronto site enrolled six healthy controls. Lastly, there were 150 healthy controls from London who completed the SDMT and PASAT.[45] Together, there was a grand total of 330 healthy control participants, although the number completing each individual MACFIMS test varied as indicated.

The inclusion criteria were consistent across the contributing studies. In order to be included, participants had to be healthy, between 18 and 65 years of age (18 and 59 for the London sample), and fluent in English (according to both participant self-assessment and examiner assessment). Participants were excluded if they had any neurological/medical/psychiatric conditions that might impede cognition, including prior head trauma, learning disability, attention-deficit disorder, mild cognitive impairment, dementia, or substance abuse. Other exclusion criteria included a history of seizures, uncorrected visual acuity problems, and current use of drugs (legal or illegal) that might have an impact on cognitive function.

Neuropsychological Measures

The MACFIMS[5,14] was administered across the four different studies by trained assistants and students under the supervision of either a neuropsychologist (LW), a neuropsychiatrist (AF), or a neurologist with expertise in cognition in MS (SM). The full MACFIMS [PASAT (3" and 2"), oral SDMT, verbal fluency (FAS and Animal Naming substituted for COWAT),[46] the JLO, CVLT–II, BVMT–R, and D–KEFS Sorting] was administered in the SUNSCREEN study. In the IPSIMS study, participants were administered all MACFIMS tests but the JLO and CVLT–II. In the BICAMS study, participants were administered all MACFIMS tests but D–KEFS Sorting and JLO. In addition, they received only the learning trials from the CVLT–II and the BVMT–R. The London participants completed only the SDMT and PASAT.

Data Analysis

Discrete normative data were established by first subdividing the sample according to three age groupings: 18-35, 36-50, and 51-65. Groupings were chosen to ensure the maximum number of participants in each cell, as well as to ensure that the data would be available for evaluating the performance of young, middle-aged, and older adults. Descriptive statistics (means and standard deviations) were calculated for each group on each measure.

For the regression-based norms, a series of linear regressions were calculated according to a procedure already well-established in the literature.[47] In most instances, just the summary score for the measure was employed, but for the CVLT–II and the BVMT–R, additional analyses were performed for relevant sub-scores. In each regression model, raw scores were the criterion variables, and age (in years), gender (coded as 1 = male, 2 = female), and education (in years) were the predictor variables entered as a single block (i.e., the "Enter" method). Given that multiple regression models were examined, a p value <0.01 was utilized throughout. As in Duff & Ramezani (2015),[47] for each regression model, the constant and non-standardized coefficients were used to generate normative formulae for ease of clinical use.

Results

See Table 1 for demographic information broken down by individual cognitive tests. In addition, for further reference, Supplementary Table 1 lists demographics for the full sample broken down by study/location. In the final sample of 330 participants, the mean age was 39.78 (11.78) years, most were female (74 vs. 26% male) given that they were selected to match the female-dominant MS populations in their respective studies, and the mean level of education was equivalent to a community college education (~15 years). Mean scores on the MACFIMS tasks fell generally within the average range when compared to published normative data for each test, suggesting that our sample was representative of the larger population.

Discrete Norms

See Table 2 for age-based discrete normative data.

Regression-Based Norms

See Table 3 for individual regression-based formulae associated with each MACFIMS test. Demographic variables varied in their contribution to each MACFIMS test in the regression models, predicting 0-18% of the variance, with a mean of 7% of the variance in the MACFIMS test scores being explained by these demographic variables. The only regression equations that failed to reach statistical significance were for the FAS and Animal Fluency (although they approached significance) and the CVLT–II, primarily because in each of these equations demographics accounted for less than 4% of the variance (on the primary measures).

The following tests and subtests were more influenced by demographic variables as measured with standardized β weights (see Supplementary Table 2) greater than 0.20 (specific demographic variable listed in parentheses): SDMT (age); 2s PASAT (education); JLO (sex, age); CVLT–II Free Recall (education); CVLT–II List B Free Recall (education); CVLT–II Short Delay Free Recall (age, education); CVLT–II Long Delay Free Recall (age, education); CVLT–II Long Delay Cued Recall (age); CVLT–II, Total Intrusions (age, education); CVLT–II, Total Repetitions (age, education); BVMT–R Total Recall (age, education); BVMT–R Delay Recall (age); D–KEFS Confirmed Correct Sorts (sex); and D–KEFS Free Sort Description Score (sex).

Discussion

This project aimed to establish Canadian discrete and regression-based normative data for the MACFIMS battery within a partially co-normed framework. Although there were four studies contributing to the healthy control data derived here, and thus not completely co-normed, the majority of tests in the battery were completed by more than 100 participants. Although regression-based data are available for the MACFIMS derived

https://doi.org/10.1017/cjn.2017.199 Published online by Cambridge University Press

Table 1: MACFIMS tests and associated sample sizes, age and education

MACFIMS test	*n*	Sex, M/F *n* (%)	Age in years, mean (*SD*)	Education in years, mean (*SD*)
SDMT	328	M = 84 (25.6)	39.66 (11.71)	15.22 (2.27)
		F = 244 (74.4)	Range = 18-65	Range = 11-25
2s PASAT	178	M = 37 (20.79)	41.53 (12.18)	15.56 (2.13)
		F = 141 (79.21)	Range = 20-65	Range = 12-21
3s PASAT	328	M = 84 (25.61)	39.83 (11.77)	15.24 (2.29)
		F = 244 (74.39)	Range = 18-65	Range = 11-25
FAS	180	M = 37 (20.56)	41.53 (12.16)	15.55 (2.13)
		F = 143 (79.44)	Range = 20-65	Range = 12-21
Animals	179	M = 37 (20.67)	41.52 (12.19)	15.56 (2.13)
		F = 142 (79.33)	Range = 20-65	Range = 12-21
JOLO	57	M = 17 (29.82)	42.24 (13.78)	15.44 (2.22)
		F = 40 (70.18)	Range = 22-65	Range = 12-21
CVLT–II Free Recall	108	M = 24 (22.22)	42.09 (12.40)	15.85 (2.20)
		F = 84 (77.78)	Range = 20-65	Range = 12-21
CVLT–II List B Free Recall	57	M = 17 (29.82)	42.24 (13.78)	15.44 (2.22)
		F = 40 (70.18)	Range = 22-65	Range = 12-21
CVLT–II SD Free	57	M = 17 (29.82)	42.24 (13.78)	15.44 (2.22)
		F = 40 (70.18)	Range = 22-65	Range = 12-21
CVLT–II SD Cued	57	M = 17 (29.82)	42.24 (13.78)	15.44 (2.22)
		F = 40 (70.18)	Range = 22-65	Range = 12-21
CVLT–II LD Free	57	M = 17 (29.82)	42.24 (13.78)	15.44 (2.22)
		F = 40 (70.18)	Range = 22-65	Range = 12-21
CVLT–II LD Cued	57	M = 17 (29.82)	42.24 (13.78)	15.44 (2.22)
		F = 40 (70.18)	Range = 22-65	Range = 12-21
CVLT–II, Total Intrusions	57	M = 17 (29.82)	42.24 (13.78)	15.44 (2.22)
		F = 40 (70.18)	Range = 22-65	Range = 12-21
CVLT–II, Total Repetitions	57	M = 17 (29.82)	42.24 (13.78)	15.44 (2.22)
		F = 40 (70.18)	Range = 22-65	Range = 12-21
BVMT–R, Total Recall	180	M = 37 (20.56)	41.53 (12.16)	15.55 (2.13)
		F = 143 (79.44)	Range = 20-65	Range = 12-21
BVMT–R, Learning	180	M = 37 (20.56)	41.53 (12.16)	15.55 (2.13)
		F = 143 (79.44)	Range = 20-65	Range = 12-21
BVMT–R, Delay Recall	129	M = 30 (23.26)	41.38 (12.70)	15.25 (2.06)
		F = 99 (76.74)	Range = 20-65	Range = 12-21
BVMT–R, Percent Retained	129	M = 30 (23.26)	41.38 (12.70)	15.25 (2.06)
		F = 99 (76.74)	Range = 20-65	Range = 12-21
D–KEFS, Confirmed Correct Sorts	129	M = 30 (23.26)	41.38 (12.70)	15.25 (2.06)
		F = 99 (76.74)	Range = 20-65	Range = 12-21
D–KEFS, Free Sort Description Score	129	M = 30 (23.26)	41.38 (12.70)	15.25 (2.06)
		F = 99 (76.74)	Range = 20-65	Range = 12-21

from an American population, research has emphasized the inadvisability of using American normative data to interpret cognitive test findings from Canadian patients, at least in part due to the social, economic, and educational differences between the two countries.[39] Thus, the current data will allow clinicians to make direct comparisons between scores on individual tests of the MACFIMS in reference to a more consistent and demographically relevant population of healthy controls than that afforded by previously available norms and by published test norms (which, again, are largely American). A case example of the application of

https://doi.org/10.1017/cjn.2017.199 Published online by Cambridge University Press

Table 2 : MACFIMS discrete norms for Canadian sample

	18-35			36-50			51-65		
MACFIMS test	***M***	***SD***	***n***	***M***	***SD***	***n***	***M***	***SD***	***n***
SDMT	66.27	9.71	131	59.73	8.7	125	56.64	8.86	72
2s PASAT	31.71	9.42	63	33.29	11.96	66	31.55	11.53	49
3s PASAT	47.92	8.07	130	47.67	9.89	124	46.43	11.91	74
FAS	41.67	10.85	64	42.63	10.76	66	44.5	9.78	50
Animals	25.73	5.52	64	23.83	3.9	65	24.12	5.36	50
JOLO	26.86	2.88	22	24.88	3.42	16	24.42	4.72	19
CVLT–II, Free Recall	58.69	7.14	36	57.65	7.85	40	56.34	9.12	32
CVLT–II, List B Free Recall	7.05	2.26	22	6.06	2.08	16	6.16	2.54	19
CVLT–II, SD Free	12.91	1.93	22	13.31	2.52	16	11.37	2.81	19
CVLT–II, SD Cued	13.5	1.82	22	13.81	2.04	16	11.42	2.27	19
CVLTII, LD Free	13.45	1.71	22	13.5	2.5	16	11.58	2.85	19
CVLT–II LD Cued	13.91	1.41	22	14.31	1.89	16	12	2.24	19
CVLT–II, Total Intrusions*	1.55	1.92	22	1.31	1.82	16	2.58	2.09	19
CVLT–II, Total Repetitions*	5.45	4.56	22	6.13	3.7	16	8.53	6.67	19
BVMT–R, Total Recall	28.75	4.8	64	28.2	5.19	66	24	6.83	50
BVMT–R, Learning	3.72	2.03	64	4.03	1.73	66	4.04	1.92	50
BVMT–R, Delay Recall	10.8	1.44	50	10.29	1.57	42	9.76	2.1	37
BVMT–R, Percent retained	96.6	5.69	50	94.18	7.51	42	98.24	4.2	37
D–KEFS, Confirmed Correct Sorts	10.44	1.92	50	10.21	1.65	42	10.1	1.49	31
D–KEFS, Free Sort Description Score	39.06	7.53	50	38.17	6.8	42	37.81	6.61	31

*Must reverse direction of sign on resulting *z* score.

Table 3: MACFIMS regression-based norms for Canadian sample

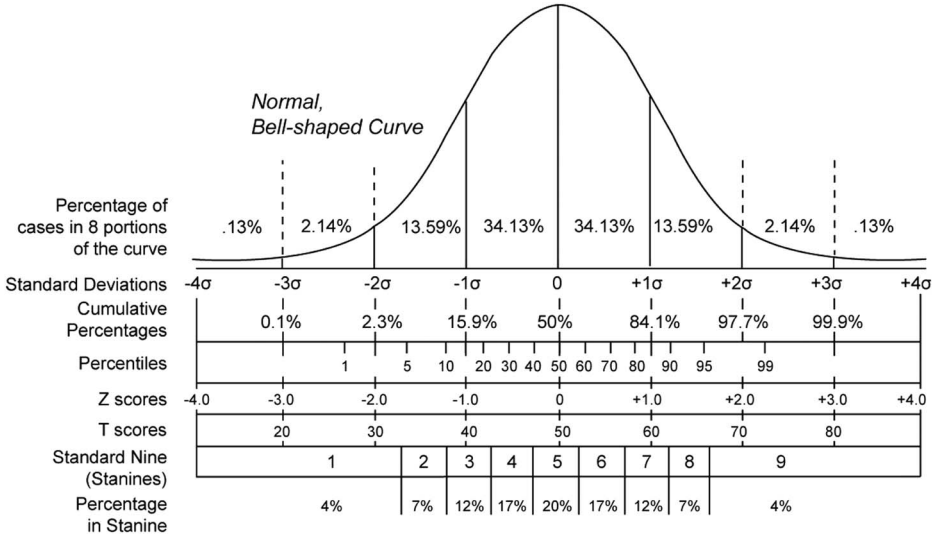
MACFIMS test	***n***	**Regression-based normative formulae**	**Adjusted R^2**	***SE* est.**	***F***
SDMT	328	70.62 + (sex × 2.83) – (age × 0.36) + (educ × 1.00)	0.18	9.03	24.09
2s PASAT	178	21.49 – (sex × 4.07) + (age × 0.02) + (educ × 1.10)	0.05	10.69	4.127
3s PASAT	328	39.24 – (sex × 1.95) – (age × 0.02) + (educ × 0.82)	0.04	9.53	5.32
FAS	180	31.86 – (sex × 0.87) + (age × 0.09) + (educ × 0.56)	0.01	10.5	1.39
Animals	179	22.53 – (sex × 0.52) – (age × 0.06) + (educ × 0.35)	0.03	4.91	2.82
JOLO	57	36.64 – (sex × 2.83) – (age × 0.08) – (educ × 0.20)	0.18	3.46	5.19
CVLT–II, Free Recall	108	48.77 – (sex × 0.33) – (age × 0.08) + (educ × 0.81)	0.04	7.84	2.6
CVLT–II, List B Free Recall	57	4.42 – (sex × 0.18) – (age × 0.02) + (educ × 0.22)	0.02	2.29	1.35
CVLT–II, SD Free	57	11.62 + (sex × 0.22) – (age × 0.06) + (educ × 0.18)	0.07	2.43	2.39
CVLT–II, SD Cued	57	12.84 + (sex × 0.62) – (age × 0.07) + (educ × 0.12)	0.13	2.12	3.72
CVLT–II, LD Free	57	11.53 + (sex × 0.27) – (age × 0.06) + (educ × 0.23)	0.11	2.35	3.37
CVLT–II, LD Cued	57	12.91 + (sex × 0.60) – (age × 0.06) + (educ × 0.12)	0.12	1.95	3.49
CVLT–II, Total Intrusions*	57	–2.42 – (sex × 0.32) + (age × 0.04) + (educ × 0.21)	0.06	1.93	2.28
CVLT–II, Total Repetitions*	57	–1.49 + (sex × 0.66) + (age × 0.08) + (educ × 0.25)	0	5.25	1.01
BVMT–R, Total Recall	180	29.21 – (sex × 0.30) – (age × 0.16) + (educ × 0.34)	0.12	5.54	8.86
BVMT–R, Learning	180	4.89 – (sex × 0.43) + (age × 0.01) – (educ × 0.04)	0	1.89	0.77
BVMT–R, Delay Recall	129	11.89 – (sex × 0.39) – (age × 0.04) + (educ × 0.05)	0.08	1.66	4.67
BVMT–R, Percent Retained	129	91.72 + (sex × 0.30) + (age × 0.03) + (educ × 0.18)	–0.02	6.21	0.34
D–KEFS, Confirmed Correct Sorts	129	11.83 – (sex × 1.23) – (age × 0.01) + (educ × 0.07)	0.07	1.93	4.12
D–KEFS, Free Sort Description Score	129	41.67 – (sex × 4.97) – (age × 0.06) + (educ × 0.49)	0.08	7.98	4.49

Sex: 1 = male, 2 = female; age: enter age in years; educ: enter education in years.
*Must reverse direction of sign on resulting *z* score.

https://doi.org/10.1017/cjn.2017.199 Published online by Cambridge University Press

Text Box 1: Case example of the application of regression-based normative data

Presume that one has assessed a 43-year-old woman with secondary-progressive MS who has 14 years of education. She has obtained a raw score of 21 on the BVMT–R Total Recall index. Normative data from the published BVMT–R manual yields a z score of –0.82. Using the current discrete norms for the 36 to 50 age group, her z score would be –1.39. Using the regression equation in Table 3, her demographically corrected predicted score would be 26.49 (i.e., 29.21 – [sex × 0.30] – [age × 0.16] + [educ × 0.34] = 29.21 – [2 × 0.30] – [43 × 0.16] + [14 x 0.34] = 26.49). The predicted score is subtracted from the observed score, and the result is divided by the standard error of the estimate for that subtest (i.e., [21 – 26.49]/5.54). This yields a z score of –0.99. As expected, given findings with the Wechsler scales,[39] Canadian raw scores on the BVMT–R are generally slightly higher than American normative data. Thus, when using the current Canadian normative data to calculate a standard score, this patient is more severely impaired than the American data would suggest. Although whether the current discrete or regression-based norms are harsher depends on the task, both forms of the Canadian norms suggest more impaired functioning for this particular individual on other tasks as well (e.g., SDMT: published $z = –2.05$; discrete $z = –2.61$; regression $z = –4.19$; FAS: published $z = –2.0$; discrete $z = –2.10$; regression $z = –3.94$).



http://psychology.wikia.com/wiki/Bell_curve_grading

regression-based normative data is presented in Text Box 1. Age, education, and sex predicted 0 to 18% of the variance in MACFIMS scores, with a mean of 7%. Age and education had the most frequent impact. The impact of demographics is small but relatively consistent with that noted in other regression-based normative data. The American regression-based MACFIMS normative study found that age, education, and sex predicted 7-28% of the variance, with age having the most frequent impact.[37] Similar data for the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) in older adults revealed that age, education, sex, and race predicted 5-28% of the variance in scores.[47] Demographics explained from about 10 to 30% of the variance in scores on the Halstead–Reitan battery.[48,49] Thus, the minimal but variable range in demographic influence observed currently is consistent with past literature.

As can be expected, the influence of particular demographic variables varied between MACFIMS tests. Tasks measuring processing speed and working memory were affected to varying degrees by demographics. SDMT performance was influenced mostly by age, whereas PASAT performance was more significantly impacted by education. The age effect on SDMT performance has been demonstrated in the literature,[50] and the lack of age effect on the PASAT is not unexpected given that the literature demonstrates a more inconsistent influence of age.[51] Education has been demonstrated to impact PASAT performance in healthy individuals (traditional scoring)[52] and in people with MS (as mediated by mathematical ability).[53] The latter is contributing to the growing literature demonstrating that the PASAT is not likely an appropriate instrument for monitoring cognition in MS despite the entrenchment of this measurement tool in the MS literature.[54,55] The lack of demographic influence on the COWAT is somewhat surprising, given that the literature demonstrates a general decline in performance with age and better scores in more highly educated individuals.[25] The lack of age effects in our 18-to-65-year-old sample may be due to the fact that the greatest influence of age is seen at the extremes (i.e., childhood and old age).[25] The lack of influence of education in our sample is likely due to the fact that the vast majority of our subjects had a high school education or above. The JLO was most influenced by age and sex, as has also been demonstrated in the literature.[10] The verbal memory test (CVLT–II) was most influenced by age and education, consistent with the findings of the test authors.[11] Nonetheless, the current results did not yield the female advantage noted by the test authors,[11] perhaps due to the fact that the current gender distribution was biased toward females. Performance on the nonverbal memory test (BVMT–R) was similar to that on the verbal task, with age and education being most contributory. The test author reported that age is

https://doi.org/10.1017/cjn.2017.199 Published online by Cambridge University Press

moderately correlated with performance, whereas the relationship of performance to education is generally weak.[56] Finally, the Sorting Test was influenced most significantly by sex. Little information is available on demographic influences on the D–KEFS measures, with the manual reporting only age effects,[13] so that it is unclear whether sex is also a variable that typically influences performance.

One limitation of our study is that some of the tasks (i.e., the CVLT–II and JLO) had small sample sizes ($n = 57$) in the regression-based formulae. This value falls below the 100 participants recommended to obtain precise classifications of performance for individuals using regression-based normative data and suggests that, for these tasks in particular, caution must be used when interpreting the performance of individuals who fall at the extremes of the demographic variables of interest.[30] Indeed, it has been suggested that when numbers are smaller the norms can lead to higher rates of false negatives (or missed diagnoses), particularly in older adults and those with lower levels of education.[57] Sample size was even smaller for these same tasks in the discrete norms (e.g., as low as 16 participants in the 36-50 age group), so that, again, caution must be applied when interpreting performance. In particular, the current findings should not be utilized in individuals over the age of 65. Similarly, care should be taken when interpreting the performance of those with less than a high school education, as all but two individuals in this database had a high school education or greater. Notably, the Conference Board of Canada reports that, as of March of 2013, 88% of working-age Canadians have graduated from high school. As such, the educational background of the current sample is fairly representative of the general population. Further limitations include a lack of ethnicity data for the full sample and a lack of information on the percentage of bilingual or multilingual speakers. Although some may consider the larger proportion of females represented in this sample as a limitation, it is important to note that the data will largely be used to interpret the performance of individuals with MS. Given that the majority of people living with MS are female, the current demographic distribution is considered appropriate.

In conclusion, provision of these Canadian regression-based formulae will allow for more accurate interpretation of MACFIMS scores by allowing clinicians to correct for relevant demographic variables simultaneously, leading to improved clinical decision making for Canadian individuals with multiple sclerosis. The provision of both discrete and regression-based options allows clinicians the freedom to choose the scoring method best suited to their own practice. Our group plans to follow up on this preliminary work by more thoroughly evaluating the utility of these Canadian regression-based normative data compared to the American data in interpreting the performance of Canadian individuals with MS.

Acknowledgements

The authors wish to thank all of the research participants who graciously gave of their time and effort to complete the various studies that made up this normative database. The authors would also like to thank the funding agencies that generously provided support for these studies, including the Multiple Sclerosis Society of Canada (Drs. Walker and Feinstein), the University of Ottawa Brain and Mind Research Institute (Drs. Walker and Cousineau), and Novartis Canada (Dr. Morrow).

Statement of Authorship

Dr. Lisa Walker: study conception, Ottawa data collection, data compilation between sites, supervision of data management, data analysis/interpretation, and manuscript preparation.

Dr. David Marino: data management, data analysis/interpretation, and manuscript editing.

Jason Berard: data management, data analysis/interpretation, and manuscript editing.

Dr. Anthony Feinstein: Toronto data collection and manuscript editing.

Dr. Sarah Morrow: London data collection and manuscript editing.

Dr. Denis Cousineau: data analysis/interpretation and manuscript editing.

Disclosures

Lisa Walker reports grants from the Multiple Sclerosis Society of Canada, grants from the University of Ottawa Brain and Mind Research Institute, and grants from Novartis Canada, during the conduct of the study; other from Serono Canada and other from Novartis Canada, outside the submitted work.

Anthony Feinstein reports grants from the MS Society of Canada, personal fees from Biogen, personal fees from Novartis, personal fees from Merck–Serono, personal fees from Sanofi–Genzyme, personal fees from Roche, and personal fees from Teva, outside the submitted work.

Sarah Morrow has received honoraria for speaking, consulting, and advisory board participation from Biogen Idec, EMD Serono, Genzyme, Novartis, and Teva Neuroscience. She has acted as site principal investigator for clinical trials for Biogen Idec. She has also received investigator-initiated study funds from Biogen Idec, Genzyme, and Novartis.

David Marino, Jason Berard, and Denis Cousineau hereby declare that they have nothing to disclose.

Supplementary material

To view supplementary materials for this article, please visit https://doi.org/10.1017/cjn.2017.199

References

1. Rao SM, Leo G, Bernardin L, Unverzagt F. Cognitive dysfunction in multiple sclerosis: frequency, patterns, and predictions. Neurology. 1991;41(5):685-91.
2. Beatty WW, Paul RH, Wilbanks LL, Hames KA, Blanco CR, Goodkin DE. Identifying multiple sclerosis patients with mild or global cognitive impairment using the Screening Examination for Cognitive Impairment (SEFCI). Neurology. 1995;45(4):718-23.
3. Basso MR, Beason-Hazen S, Lynn J, Rammohan K, Bornstein RA. Screening for cognitive dysfunction in multiple sclerosis. Arch Neurol. 1996;53(10):980-4.
4. Langdon DW, Amato MP, Boringa J, et al. Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). Mult Scler. 2012;18(6):891-8; Epub ahead of print Dec 21, 2011. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3546642/.
5. Benedict RH, Fischer JS, Archibald CJ, et al. Minimal neuropsychological assessment of MS patients: a consensus approach. Clin Neuropsychol. 2002;16(3):381-97.
6. Rao SM, Leo GJ, Ellington L, Nauertz T, Bernardin L, Unveragt F. Cognitive dysfunction in multiple sclerosis, II: impact on employment and social functioning. Neurology. 1991;41(5):692-6.

https://doi.org/10.1017/cjn.2017.199 Published online by Cambridge University Press

7. Gronwall DM. Paced auditory serial addition task: a measure of recovery from concussion. Percept Mot Skills. 1977;44(2): 367-73.
8. Smith A. Symbol Digit Modalities Test. Los Angeles: Western Psychological Services; 1982.
9. Benton AL, Hamsher K. Multilingual Aphasia Examination. Iowa City: AJA Associates; 1989.
10. Benton AL, Sivan AB, Hamsher K, Varney NR, Spreen O. Contributions to Neuropsychological Assessment, 2nd ed. New York: Oxford University Press; 1994.
11. Delis DC, Kramer JH, Kaplan E, Ober BA.. California Verbal Learning Test, 2nd ed. San Antonio, TX: Psychological Corporation; 2000.
12. Benedict RH. The Brief Visual Memory Test–Revised. Lutz, FL: Psychological Assessment Resources; 1997.
13. Delis DC, Kaplan E, Dramer JH. Delis–Kaplan Executive Function System. San Antonio, TX: The Psychological Corporation; 2001.
14. Benedict RH, Cookfair D, Gavett R, et al. Validity of the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS). J Int Neuropsychol Soc. 2006;12(4):549-58.
15. Strober L, Englert J, Munschauer F, Weinstock-Guttman B, Rao SM, Benedict RH. Sensitivity of conventional memory tests in multiple sclerosis: comparing the Rao Brief Repeatable Neuropsychological Battery and the Minimal Assessment of Cognitive Function in MS. Mult Scler. 2009;15(9):1077-84.
16. Hoogs M, Kaur S, Smerbeck A, Weinstock-Guttman B, Benedict RH. Cognition and physical disability in predicting health-related quality of life in multiple sclerosis. Int J MS Care. 2011;13(2):57-63; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3882954/pdf/i1537-2073-13-2-57.pdf.
17. Benedict RH, Zivadinov R. Reliability and validity of neuropsychological screening and assessment strategies in MS. J Neurol. 2007;254(Suppl 2):II22-5.
18. Hanna J, Feinstein A, Morrow SA. The association of pathological laughing and crying and cognitive impairment in multiple sclerosis. J Neurol Sci. 2016;361:200-3; Epub ahead of print Jan 5.
19. Morrow SA, Drake A, Zivadinov R, Munschauer F, Weinstock-Guttman B, Benedict RH. Predicting loss of employment over three years in multiple sclerosis: clinically meaningful cognitive decline. Clin Neuropsychol. 2010;24(7):1131-45.
20. van Gorp WG, Satz P, Mitrushina M. Neuropsychological processes associated with normal aging. Dev Neuropsychol. 1990;6(4): 279-90.
21. Heaton RK, Ryan L, Grant I. Demographic influences and use of demographically corrected norms in neuropsychological assessment. In Grant I, Adams KM editors. Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders, 3rd ed. New York: Oxford University Press; 2009, P. 127-55.
22. Vanderploeg RD, Axelrod BN, Sherer M, Scott J, Adams RL. The importance of demographic adjustments on neuropsychological test performance: a response to Reitan and Wolfson. Clin Neuropsychol. 1997;11(2):210-7.
23. Lezak MD, Howieson DB, Bigler ED, Tranel D. Neuropsychological Assessment, 5th ed. New York: Oxford University Press; 2012.
24. Mitrushina M, Boone KB, Razani J, D'Elia LF. Handbook of Normative Data for Neuropsychological Assessment, 2nd ed. New York: Oxford University Press; 2005.
25. Strauss E, Sherman EM, Spreen O. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, 3rd ed. New York: Oxford University Press; 2006.
26. Crawford JR, Howell DC. Regression equations in clinical neuropsychology: an evaluation of statistical methods for comparing predicted and obtained scores. J Clin Exp Neuropsychol. 1998; 20(5):755-62.
27. Testa SM, Winicki JM, Pearlson GD, Gordon B, Schretlen DJ. Accounting for estimated IQ in neuropsychological test performance with regression-based techniques. J Int Neuropsychol Soc. 2009;15(6):1012-22; Epub ahead of print Oct 2.
28. Bergman I, Almkvist O. Neuropsychological test norms controlled for physical health: does it matter? Scand J Psychol. 2015; 56(2):140-50; Epub ahead of print Oct 30, 2014.
29. Oosterhuis HE, van der Ark LA, Sijtsma K. Sample size requirements for traditional and regression-based norms. Assessment. 2016;23(2):191-202; Epub ahead of print May 4, 2015.
30. Berrigan LI, Fisk JD, Walker LA, et al. Reliability of regression-based normative data for the oral Symbol Digit Modalities Test: an evaluation of demographic influences, construct validity, and impairment classification rates in multiple sclerosis samples. Clin Neuropsychol. 2014;28(2):281-99; Epub ahead of print Jan 20.
31. Eshaghi A, Riyahi-Alam S, Roostaei T, et al. Validity and reliability of a Persian translation of the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS). Clin Neuropsychol. 2012;26(6):975-84; Epub ahead of print Jun 11.
32. Dusankova JB, Kalincik T, Havrdova E, Benedict RH. Cross-cultural validation of the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS) and the Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). Clin Neuropsychol. 2012;26(7):1186-200; Epub ahead of print Oct 4.
33. Rohling ML, Miller RM, Axelrod BN, Wall JR, Lee AJH, Kinikini DT. Is co-norming required? Arch Clin Neuropsychol. 2015;30(7): 611-33; Epub ahead of print Jul 6.
34. Russell EW, Russell SLK, Hill BD. The fundamental psychometric status of neuropsychological batteries. Arch Clin Neuropsychol. 2005;20(6):785-94.
35. Smerbeck AM, Parrish J, Yeh EA, et al. Regression-based pediatric norms for the Brief Visuospatial Memory Test–Revised and the Symbol Digit Modalities Test. Clin Neuropsychol. 2011; 25(3):402-12; Epub ahead of print Mar 7.
36. Smerbeck AM, Parrish J, Yeh EA, et al. Regression-based norms improve the sensitivity of the National MS Society Consensus Neuropsychological Battery for Pediatric Multiple Sclerosis (NBPMS). Clin Neuropsychol. 2012;26(6):985-1002.
37. Parmenter BA, Testa SM, Schretlen DJ, Weinstock-Guttman B, Benedict RH. The utility of regression-based norms in interpreting the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS). J Int Neuropsychol Soc. 2010;16(1):6-16.
38. Miller JL, Weiss LG, Beal AL, Saklofske DH, Zhu J, Holdnack JA. Intelligent use of intelligence tests: empirical and clinical support for Canadian WAIS–IV norms. J Psychoeduc Assess. 2015;33(4):312-28.
39. Harrison AG, Armstrong IT, Harrison LE, Lange RT, Iverson GL. Comparing Canadian and American normative scores on the Wechsler Adult Intelligence Scale, 4th edition. Arch Clin Neuropsychol. 2014;29(8):737-46; Epub ahead of print Oct 13. Available from: https://academic.oup.com/acn/article-lookup/doi/10.1093/arclin/acu048.
40. Longman RS, Saklofske DH, Fung TS. WAIS–III percentile scores by education and sex for U.S. and Canadian populations. Assessment. 2007;14(4):426-32.
41. Geidraitiene N, Kizlatiene R, Kaubrys G. The BICAMS Battery for Assessment of Lithuanian-Speaking Multiple Sclerosis Patients: relationship with age, education, disease disability, and duration. Med Sci Monit. 2015;21:3853-9; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4678925/pdf/medscimonit-21-3853.pdf.
42. O'Connell K, Langdon D, Tubridy N, Hutchinson M, McGuigan C. A preliminary validation of the Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS) tool in an Irish population with multiple sclerosis (MS). Mult Scler Relat Disord. 2015;4(6):521-5; Epub ahead of print Aug 7.
43. Sandi D, Rudisch T, Fuvesi J, et al. The Hungarian validation of the Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS) battery and the correlation of cognitive impairment with fatigue and quality of life. Mult Scler Relat Disord. 2015; 4(6):499-504; Epub ahead of print Jul 8.
44. Spedo CT, Frndak SE, Marques VD, et al. Cross-cultural adaptation, reliability, and validity of the BICAMS in Brazil. Clin Neuropsychol. 2015;29(6):836-46; Epub ahead of print Oct 29.
45. Morrow SA. Normative data for the Stroop Color Word Test for a North American population. Can J Neurol Sci. 2013; 40(6):842-7.
46. Tombaugh TN, Kozak J, Rees L. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Arch Clin Neuropsychol. 1999;14(2): 167-77.

https://doi.org/10.1017/cjn.2017.199 Published online by Cambridge University Press

47. Duff K, Ramezani A. Regression-based normative formulae for the Repeatable Battery for the Assessment of Neuropsychological Status for older adults. Arch Clin Neuropsychol. 2015;30(7): 600-4; Epub ahead of print Aug 18. Available from: https://academic.oup.com/acn/article-lookup/doi/10.1093/arclin/acv052.
48. Heaton RK, Miller SW, Taylor MJ, Grant I. Revised Comprehensive Norms for an Expanded Halstead–Reitan Battery: Demographically Adjusted Neuropsychological Norms for African-American and Caucasian Adults. Lutz, FL: Psychological Assessment Resources; 2004.
49. Moses JA, Pritchard DA, Adams RL. Normative corrections for the Halstead–Reitan Neuropsychological Battery. Arch Clin Neuropsychol. 1999;14(5):445-54.
50. Selnes OA, Jacobson L, Machado AM, et al. Normative data for a brief neuropsychological screening battery. Percept Mot Skills. 1991;73(2):539-50.
51. Tombaugh TN. A comprehensive review of the Paced Auditory Serial Addition Test. Arch Clin Neuropsychol. 2006;21(1):53-76; Epub ahead of print Nov 14, 2005. Available from: http://www.sciencedirect.com/science/article/pii/S0887617705001320.
52. Gonzalez R, Grant I, Miller SW, et al. Demographically adjusted normative standards for new indices of performance on the Paced Auditory Serial Addition Task (PASAT). Clin Neuropsychol. 2006;20(3):396-413.
53. Sandry J, Paxton J, Sumowski JF. General mathematical ability predicts PASAT performance in MS patients: implications for clinical interpretation and cognitive reserve. J Int Neuropsychol Soc. 2016;22(3):375-8; Epub ahead of print Jan 29.
54. Sonder JM, Burggraaff J, Knol DL, Polman CH, Uitdehaag BMJ. Comparing long-term results of PASAT and SDMT scores in relation to neuropsychological testing in multiple sclerosis. Mult Scler.. 2014;20(4):481-8; Epub ahead of print Sep 9, 2013.
55. Walker LA, Cheng A, Berard J, Berrigan LI, Rees LM, Freedman MS. Tests of information processing speed: what do people with multiple sclerosis think about them? Int J MS Care. 2012; 14(2):92-9; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3883004/pdf/i1537-2073-14-2-92.pdf.
56. Benedict RHB, Schretlen DJ, Groninger L, Dobraski M, Shpritz B. Revision of the Brief Visuospatial Memory Test: studies of normal performance, reliability, and validity. Psychol Assess. 1996; 8(2):145-53.
57. Fastenau PS. Validity of regression-based norms: an empirical test of the comprehensive norms with older adults. J Clin Exp Neuropsychol. 1998;20(6):906-16.

https://doi.org/10.1017/cjn.2017.199 Published online by Cambridge University Press