BRIEF COMMUNICATIONS

Influence of recruitment and participation bias in neuropsychological research among MS patients

J. AUBREY DUQUIN,[1] BRETT A. PARMENTER,[2] AND RALPH H.B. BENEDICT[1]

[1]Department of Neurology, Division of Cognitive and Behavioral Neurosciences, SUNY at Buffalo School of Medicine and Biomedical Sciences, Buffalo, New York
[2]Department of Psychology, Washington State University, Pullman, Washington

Abstract

The potential influence of recruitment context in neuropsychological (NP) research is seldom addressed in the literature. Our experience with a previous large-scale study of MS cognitive impairment led us to speculate that referral questions and motivation bias significantly impact conclusions drawn from NP testing. We re-analyzed data from Benedict et al. (2006) and compared the results obtained across three groups of patients: paid research volunteers, clinical patients undergoing evaluation for routine monitoring of cognitive status, and clinical patients referred for diagnostic clarification, determination of disability benefits, and other more complex issues. Research volunteers were significantly younger and less frequently cognitively impaired compared to the clinical groups, greatly affecting prevalence estimates (45.6 to 65.6%). Significant correlations between depression and cognitive impairment were present only among the research volunteers. NP testing significantly predicted vocational outcomes in all groups. These results suggest that greater attention should be paid to the reasons patients agree to participate in NP testing. (*JINS*, 2008, *14*, 494–498.)

Keywords: Cognition, Depression, Motivation, Incentive, Selection, Generalizability

INTRODUCTION

The generalizability of behavioral research depends on random recruitment and freedom from selection bias. However, neuropsychological (NP) testing is an uncommon medical procedure that can be arduous and time consuming. If patients are not adequately reimbursed for their time, their motivation to participate may stem from idiosyncratic characteristics, thus creating a selection bias. In the multiple sclerosis (MS) literature, the recruitment context in NP research is often poorly described. Nelson and colleagues (1998) compared clinical and demographic measures between two MS groups: (a) those treated at a center of specialty care of MS and (b) patients randomly recruited from the surrounding community and treated by a non-MS-specialist neurologist. Patients treated at the center were

younger, had more mobility impairment, had earlier diagnosis supported by laboratory tests, and more often reported recent disease worsening. The authors cautioned researchers that limiting research to MS center patients could lead to erroneous conclusions pertaining to demographics, disease course and clinical presentation. Another study, conducted in the university setting, investigated methods of recruiting healthy volunteers (Tomporowski et al., 1993). Cognitive tests obtained from university psychology students required to participate for class credit were compared to non-psychology students paid to undergo testing. The latter group performed better on tests of attention and memory. These studies suggest that differences in recruitment strategy and incentives for participation may seriously impact the conclusions drawn from NP research.

During the course of a recent cross-sectional study of 291 MS patients (Benedict et al., 2006), we found that seven NP tests discriminated MS patients from healthy controls, and relapsing-remitting (RR) from secondary-progressive (SP) patients. Tests that emphasized episodic memory and

processing speed were most sensitive and logistic regression models revealed that tests emphasizing processing speed, verbal memory, and executive function were predictive of vocational status. Depression, was not significantly related to cognitive performance, consistent with some previous findings (Benedict et al., 2002a) but not others (Arnett et al., 1999; Arnett et al., 2002). Patients were gleaned from a number of different sources as documented prior to testing. Therefore, we were able to retrospectively categorize patients into groups based on their recruitment context and incentive to participate. We decided to investigate whether such bias may affect conclusions drawn about commonly asked research questions in the MS literature concerning the frequency of cognitive impairment, correlation with depression, and external validity.

METHOD

The data from Benedict et al. (2006), including 291 patients diagnosed with clinically definite MS and 56 normal controls, were reanalyzed. Exclusionary criteria were (a) medical disorder other than MS affecting cognitive function, (b) psychiatric disorder (American Psychiatric Association, APA, 2000) other than mood, personality, or behavioral change following the onset of MS, (c) drug or alcohol dependence or current abuse, (d) motor or sensory impairment that could compromise testing, or (e) relapse or corticosteroid treatment within four weeks of assessment. Prior to participation, all research participants signed consent forms approved by institutional review boards.

Patients were coded on a new variable called "recruitment context." The Research Volunteer (RV) group included patients volunteering to undergo NP testing for financial compensation ($n = 57$); the Routine Monitoring (RM) group consisted of clinical patients referred for evaluation in order to monitor for changes in cognitive capacity ($n = 106$), and patients referred for evaluation of a specified clinical problem [repeated complaints of impairment ($n = 64$), alleged failure in work place ($n = 31$), differential diagnosis or psychiatric co-morbidity ($n = 33$)] comprised the Clinically Complex (CC) group ($n = 128$).

The MACFIMS (Minimal Assessment of Cognitive Function in MS) neuropsychological test battery was employed, as recommended by a consensus panel (Benedict et al., 2002b) and recently validated in MS (Benedict et al., 2006). The Controlled Oral Word Association Test (COWAT) (Benton & Hamsher, 1989) and the Judgment of Line Orientation Test (JLO) (Benton et al., 1994) were used to assess language and spatial processing abilities, respectively. We used two indices from each of two memory tests, the Total Learning (TL) and Delayed Recall (DR) scores. The California Verbal Learning Test—second edition (CVLT2) (Delis et al., 2000) was employed for auditory/verbal memory and the Brief Visuospatial Memory Test–Revised (BVMTR) (Benedict, 1997) was used for visual/spatial memory. Rao adaptations (Rao et al., 1991a) of the Symbol Digit Modalities Test (SDMT) (Smith, 1982) and the Paced Auditory

Serial Addition Test (PASAT) (Gronwall, 1977) were used to assess processing speed and working memory. Number correct on the 3.0 and 2.0 ISI PASAT were added together and coded as a single score. The Delis-Kaplan Executive Function System Sorting Test (DKEFS) (Delis et al., 2001) was employed for the assessment of executive function, including the Correct Sorts and Descriptions scores. Finally, the Beck Depression Inventory—Fast Screen (BDIFS) (Beck et al., 2000), validated in the MS population (Benedict et al., 2003), was included as an index of depression.

As in our previous work (Benedict et al., 2006), patients were classified into two groups based on their employment/disability status. Working patients were employed at least 20 hours per week. Disabled patients were receiving formal disability benefits from either public or private sources or were unemployed for reasons reported by them or informants to be disease-related. The classification was determined prior to NP testing.

The analysis plan was to examine common research questions in the MS literature using conventional statistical techniques and determine if the conclusions reached would differ across the motivation groups (RV, RM, CC). We employed an alpha criterion of $p < .05$ for statistical significance throughout. ANOVA and Chi-square tests were utilized to compare disease characteristics. Effect sizes were calculated using Cohen's d statistic. Following our previous work (Benedict et al., 2006), z-scores were calculated for each individual NP test based on a demographically matched sample of 56 healthy control volunteers. Patients were classified as impaired if they performed in the impaired range (≤ 1.5 standard deviations below normal controls) on two or more NP tests. Age adjusted z-scores were also calculated. Using the control group, we regressed raw scores on age and saved the standardized residuals for each NP measure. We then applied the regression equations from the healthy controls to calculate age predicted raw scores for each MS patient. These predicted scores were subtracted from the obtained scores and the differences were divided by the standard deviation of the control group's residuals.

In the first analysis we asked the question, what is the frequency of cognitive impairment in MS? To answer the question we used chi-square tests to determine if the proportion of cognitively-impaired patients differed across group.

Next, we asked, is cognitive impairment associated with depression in MS? The frequency of depression was compared statistically across groups using Chi-square tests. Partial correlations were examined between BDIFS and each cognitive measure controlling for age.

Finally, ANOVA was used to compare differences in mean performance on NP tests between two vocational subgroups of patients (employed vs. disabled). In the Benedict et al., 2006 paper, tests emphasizing memory, processing speed, and executive function were most valid for discriminating MS patients from healthy controls, and were most predictive of vocational status. Thus, we limited our analysis to tests in these domains to address this question.

RESULTS

RV patients were significantly younger (mean age $\pm$ SD; RV = 42.4 $\pm$ 8.5, RM = 46.8 $\pm$ 9.3, CC = 45.6 $\pm$ 8.6) than those in the RM group [$F(3,343) = 3.53, p = .015$]. The three groups did not differ in educational attainment, gender, or ethnicity. Expanded Disability Status Scale (Kurtzke, 1983) scores were also similar across group: RV 2.7 $\pm$ 1.9, RM 3.2 $\pm$ 1.8, CC 3.1 $\pm$ 1.8; $p =.33$. Chi-square analysis showed that the percentage of patients with progressive disease course was significantly greater in the CC group than in the RV group [$\chi^2(1) = 4.46, p = .035$]. There were no group differences in disease duration.

Analysis A, Frequency of Cognitive Impairment

Frequency of impairment was 45.6% for RV patients, 59.4% for RM and 65.6% for CC patients. Chi-square analysis showed that the percentage in the RV group was significantly lower than in the CC group [$\chi^2(1) = 6.55, p = .010$]. Frequencies of cognitive impairment for each individual test can be found in Table 1. Chi-square analysis showed that more CC patients were impaired on CVLT2-DR as compared to the RV group (21%) [$\chi^2(2) = 5.99, p = .05$] and more CC patients were impaired on the PASAT than RV patients [$\chi^2(2) = 7.42, p = .024$]. ANOVAs evaluating age-corrected z-scores revealed group differences on the PASAT [$F(2,288) = 8.52, p < .001$]. Tukey post-hoc analyses showed that the RV group performed significantly better than the RM and CC groups.

Analysis B, Correlation with Depression

As in previous work (Benedict et al., 2003) patients were considered to be depressed if their score on the BDIFS was >3 (Beck et al., 2000). ANCOVA, with age as the covariate,

comparing group means on the total score of the BDI-FS showed no significant group differences. The incidence of depression did not differ significantly between group: RV = 35.1%, RM = 47.2%, CC = 44.5%. Significant partial correlations (controlling for age) between the BDI-FS and NP tests were only revealed within the RV group: JLO ($r = -.49$, $p < .001$), BVMTR-DR ($r = -.32, p = .015$), PASAT ($r = -.33, p = .013$), SDMT ($r = -.42, p = .001$), DKEFS Sorting Test, correct sorts ($r = -.40$), $p = .002$.

Analysis C, Predicting Vocational Status

Chi-square analyses revealed significant differences [$\chi^2(2) = 6.59, p =.037$] for the proportion of disabled participants in each group. The proportion of disabled participants in the RV group (39.39%) was significantly lower than that of RM (62.11%) and CC (63.63%) patients [$\chi^2(1) = 5.13, p =.023$ and $\chi^2(1) = 6.13, p =.013$, respectively]. ANOVAs using the age-corrected cognitive data were conducted within each MS group to examine the effect of NP test on vocational status (Table 2). Within the RV group, significant differences between vocational subgroups were revealed on the SDMT [$F(1,31) = 4.85, p = .035$]. Significant differences were noted on CVLT2-TL [$F(1,93) = 8.22, p = .005$], CVLT2-DR [$F(1,93) = 8.48$, $p = .004$], SDMT [$F(1,93) = 8.11, p = .005$], DKEFS Sorting-Correct Sorts [$F(1,93) = 4.80, p =.031$], and DKEFS Sorting Test-Description Score [$F(1,93) = 7.69$, $p = .007$] within the RM group. Significant differences between groups were noted on every NP test except the COWAT and JLO in the CC group, with p-values ranging from $p = .007$ to $p < .001$.

DISCUSSION

We investigated three basic questions addressed in many previously published MS studies and determined if the results

Table 1. Neuropsychological test data per MS group

	Research Volunteers (RV)			Routine Monitoring (RM)			Clinically Complex (CC)		
	Mean	SD	% impaired	Mean	SD	% impaired	Mean	SD	% impaired
COWAT	37.5	9.3	7.1	34.6	11.0	17.0	33.3	10.8	12.5
JLO	24.4	4.2	15.8	22.3	6.6	20.8	22.2	5.9	26.6
CVLT2-TL	51.4	11.9	22.8	47.9	10.8	29.2	47.2	11.9	32.8
CVLT2-DR	10.9	3.5	21.1	9.9	3.6	36.8	9.6	3.7	38.3
BVMT-R-TL	23.0	7.3	31.6	20.9	7.6	46.2	20.3	7.8	48.4
BVMTR-DR	8.8	2.8	40.4	8.3	2.9	41.5	7.9	3.2	52.3
PASAT	41.3	11.1	15.8	33.8	11.8	26.4	32.5	13.5	33.6
SDMT	52.4	13.7	36.8	47.1	13.7	54.7	46.2	14.2	53.9
DKEFS Sorting–CS	10.2	2.5	10.5	9.6	2.4	13.2	9.2	2.8	20.3
DKEFS Sorting–DS	36.9	11.4	21.1	34.8	11.3	23.6	34.9	11.9	29.7

Abbreviations: COWAT = Controlled Oral Word Association Test; JLO = Judgment of Line Orientation Test; CVLT2 = California Verbal Learning Test, second edition; BVMT-R = Brief Visuospatial Memory Test–Revised; PASAT = Paced Auditory Serial Addition Test; SDMT = Symbol Digit Modalities Test; DKEFS Sorting—CS = Delis-Kaplan Executive Function System Sorting Test—Correct Sorts; DKEFS Sorting—DS = Delis-Kaplan Executive Function System Sorting Test—Description Score.

Table 2. MS group data comparing NP data for normal versus disabled vocational status

| | Research Volunteers (RV) | | | | | | Routine Monitoring (RM) | | | | | | Clinically Complex (CC) | | | | | |
| | Employed | | Disabled | | | | Employed | | Disabled | | | | Employed | | Disabled | | | |
	Mean	SD	Mean	SD	d	P	Mean	SD	Mean	SD	d	P	Mean	SD	Mean	SD	d	P
CVLT Total Learning	55.90	9.65	50.31	9.39	0.59	.113	51.53	10.02	45.00	11.03	0.62	.005	52.50	9.89	44.74	12.37	0.70	.001
CVLT Long Delay	12.30	2.98	10.08	3.33	0.70	.057	11.06	3.23	8.86	3.65	0.63	.004	11.15	2.71	9.07	4.05	0.61	.007
BVMT Total Learning	25.60	5.09	21.08	8.77	0.65	.122	22.36	7.09	19.37	7.66	0.41	.083	24.98	6.08	18.13	7.77	0.99	.000
BVMT Long Delay	9.55	1.61	8.00	3.89	0.56	.168	8.75	2.30	7.59	3.13	0.43	.012	10.03	2.02	6.76	3.22	1.25	.000
PASAT Total	46.30	8.92	37.96	12.94	0.75	.052	35.01	10.88	31.75	12.22	0.28	.237	38.18	10.43	29.91	14.69	0.65	.007
SDMT	59.15	10.96	48.31	14.11	0.86	.035	51.86	12.28	43.68	13.89	0.63	.005	55.28	10.51	41.09	14.43	1.14	.000
DKEFS Correct Sorts	11.05	2.16	9.46	2.60	0.67	.085	10.28	2.34	9.12	2.44	0.49	.031	10.53	1.71	8.41	3.19	0.87	.001
DKEFS Description Score	39.00	11.79	33.69	11.75	0.45	.251	39.19	9.84	32.53	11.69	0.62	.007	40.58	7.24	31.93	13.07	0.85	.001

Abbreviations: CVLT2 = California Verbal Learning Test, second edition; BVMT-R = Brief Visuospatial Memory Test–Revised; PASAT = Paced Auditory Serial Addition Test; SDMT = Symbol Digit Modalities Test; DKEFS = Delis-Kaplan Executive Function System.

would differ depending on the reasons patients were brought to the NP testing milieu. Patients were categorized into research volunteers (RV), clinical patients seen for routine monitoring of cognitive capacity (RM), and patients seen for more complex referral questions pertaining to differential diagnosis, psychiatric co-morbidity, determination of vocational capacity, and alike. We found that substantially different conclusions would be drawn about the frequency of cognitive impairment in MS, the relationship between depression and cognitive impairment, and to a lesser extent, the external validity of NP testing in MS, depending on the group studied.

Were our study limited to research volunteers, we would conclude that the frequency of cognitive impairment in MS is roughly 46%, a figure approximating the 43% figure reported in Rao et al.'s (1991a) seminal study of 100 MS patients recruited by advertisement from a US metropolitan area. Rao et al. (1991a) indicated that subjects were paid for their participation, but they did not disclose how much. In our study, patients were drawn from multiple studies and paid at a rate of roughly $20 per hour. It is important to consider this point. If research volunteers are not paid at a rate commensurate to or higher than their regular income, other factors may be motivating them to participate. NP testing requires considerable effort and patients who are not cognitively impaired may be less threatened and more willing to volunteer their time to please doctors/researchers or help in the cause of understanding their disease. We did not query each volunteer about these issues. Presumably, a mixture of incentives contributed to the composition of the RV group.

The RM group seems most representative of the MS population. These patients were referred to monitor NP status (cognitive and psychiatric) such that changes would be detected as early as possible. Indeed, this was the intent of the MACFIMS battery: a streamlined collection of tests known to be sensitive to MS associated impairment and administered in a relatively short time. Conclusions reached from the RM group data were intermediate between the RV and CC groups. For this group, the frequency of cognitive impairment was 59% and roughly $\frac{1}{2}$ of the NP tests discriminated employed from disabled sub-groups.

The relationship between depression and cognitive impairment has been addressed in many MS studies. In general, correlations between NP testing and depression are modest, sometimes significant (e.g., Arnett et al., 1999; Arnett et al., 2002), sometimes not (e.g., Benedict et al., 2002a). Our analysis provided an opportunity to learn if such variation is related to recruitment context. While there were no significant differences in the incidence of depression, partial correlations between NP testing and BDIFS, were significant only in the RV group. One explanation concerns the etiology of depression. The RV group may be composed of patients who are depressed in response to increased coping demands following the onset of cognitive impairment. In other words, these patients may have reactive depression which motivates them to seek information about their mental state. Alternatively,

these RV patients with time to volunteer for research may have less complicated disease, fewer side effects from pharmacological therapy, and less co-morbid illness. Fewer complications may facilitate the detection of cognition-depression associations in statistical analysis.

Finally, we examined the external validity of NP testing by searching for tests that distinguish employed vs. disabled patients. Technically, the same conclusion is reached within each group: there is a significant relationship between cognitive testing and work capability as shown in prior research (Benedict et al., 2006; Rao et al., 1991b). However, the effects are more consistent and stronger among the RM and CC patients. These patients are more cognitively impaired, hence the enhanced sensitivity of the analysis in these groups. Among patients in the RM group, we found that verbal memory, processing speed and higher executive function defects were significantly predictive of work disability.

In sum, this re-analysis of the Benedict et al. (2006) data highlights the importance of measuring and reporting the recruitment context and incentives to participate in NP testing. Such bias may lead to very different conclusions about fundamental questions in the neuropsychology of MS.

ACKNOWLEDGMENT

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.-TR). Washington, DC: American Psychiatric Association.

Arnett, P.A., Higginson, C.I., Voss, W.D., Bender, W.I., Wurst, J.M., & Tippin, J.M. (1999). Depression in multiple sclerosis: Relationship to working memory capacity. *Neuropsychology*, *13*, 546–556.

Arnett, P.A., Higginson, C.I., Voss, W.D., Randolph, J.J., & Grandey, A.A. (2002). Relationship between coping, cognitive dysfunction and depression in multiple sclerosis. *Clinical Neuropsychologist*, *16*, 341–355.

Beck, A.T., Steer, R.A., & Brown, G.K. (2000). *BDI-Fast Screen for Medical Patients: Manual*. San Antonio, TX: Psychological Corporation.

Benedict, R.H.B. (1997). *Brief Visuospatial Memory Test–Revised: Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Benedict, R.H.B., Bakshi, R., Simon, J.H., Priore, R., Miller, C., & Munschauer, F. (2002a). Frontal cortex atrophy predicts cog-

nitive impairment in multiple sclerosis. *Journal of Neuropsychiatry & Clinical Neurosciences, 14*, 44–51.

Benedict, R.H.B., Fischer, J.S., Archibald, C.J., Arnett, P.A., Beatty, W.W., Bobholz, J., Chelune, G., Fisk, J.D., Langdon, D.W., Caruso, L.S., Foley, F.W., LaRocca, N.G., Vowels, L., Weinstein, A., DeLuca, J., Rao, S.M., & Munschauer, F.E. (2002b). Minimal Neuropsychological Assessment of MS Patients: A Consensus Approach. *Clinical Neuropsychologist*, *16*, 381–397.

Benedict, R.H.B., Fishman, I., McClellan, M.M., Bakshi, R., & Weinstock-Guttman, B. (2003). Validity of the Beck Depression Inventory—Fast Screen in multiple sclerosis. *Multiple Sclerosis*, *9*, 393–396.

Benedict, R.H.B., Cookfair, D., Gavett, R., Gunther, M., Munschauer, F., Garg, N., & Weinstock-Guttman, B. (2006). Validity of the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS). *Journal of the International Neuropsychological Society*, *12*, 549–558.

Benton, A.L. & Hamsher, K. (1989). *Multilingual Aphasia Examination*. Iowa City, IA: AJA Associates.

Benton, A.L., Sivan, A.B., Hamsher, K., Varney, N.R., & Spreen, O. (1994). *Contributions to Neuropsychological Assessment* (2nd ed.). New York: Oxford University Press.

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan Executive Function System*. San Antonio, TX: Psychological Corporation.

Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (2000). *California Verbal Learning Test Manual: Second Edition, Adult Version*. San Antonio, TX: Psychological Corporation.

Gronwall, D.M.A. (1977). Paced Auditory Serial Addition Task: A measure of recovery from concussion. *Perceptual and Motor Skills*, *44*, 367–373.

Kurtzke, J.F. (1983). Rating neurologic impairment in multiple sclerosis: An Expanded Disability Status Scale (EDSS). *Annals of Neurology*, *13*, 227–231.

Nelson, L.M., Franklin, G.M., Hamman, R.F., Boteler, D.L., Baum, H.M., & Burks, J.S. (1988). Referral bias in multiple sclerosis research. *Journal of Clinical Epidemiology*, *41*, 187–192.

Rao, S.M., Leo, G.J., Bernardin, L., & Unveragt, F. (1991a). Cognitive dysfunction in multiple sclerosis. I. Frequency, Patterns and Prediction. *Neurology*, *41*, 685–691.

Rao, S.M., Leo, G.J., Ellington, L., Nauertz, T., Bernardin, L., & Unveragt, F. (1991b). Cognitive dysfunction in multiple sclerosis. II. Impact on employment and social functioning. *Neurology*, *41*, 692–696.

Smith, A. (1982). *Symbol Digit Modalities Test: Manual*. Los Angeles, CA: Western Psychological Services.

Tomporowski, P.D., Simpson, R.G., & Hager, L. (1993). Method of recruiting subjects and performance on cognitive tests: *American Journal of Psychology*, *106*, 499–521.