INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
NEUROPSYCHOLOGICAL DEFICITS AS PREDICTORS OF TREATMENT OUTCOME IN PATIENTS WITH TEMPOROMANDIBULAR DISORDERS

by

Márcio Lima Grossi, C.D., M.S.

A thesis is submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Dentistry
University of Toronto

© Copyright by Márcio Lima Grossi (Fall 1998)
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
ABSTRACT

TITLE: Neuropsychological Deficits as Predictors of Treatment Outcome in Patients with Temporomandibular Disorders. Thesis for the Doctor of Philosophy Degree, Fall 1998. Márcio Lima Grossi, C.D., M.S. Graduate Department of Dentistry, University of Toronto.

The primary objective of this study was to assess the utility of neuropsychological tests as predictors of treatment outcome for patients with temporomandibular disorders (TMD). Psychosocial and clinical variables were also included as confounders and secondary outcome measures. In the overall, the nrTMD did worse in both the neuropsychological and psychosocial assessment with higher memory deficits, sleep disturbances, levels of depression and fatigue, and lower energy levels; but both groups could not be distinguished from each other based on clinical variables usually assessed in the TMD examination. The California Verbal Learning Test - correct responses (CVLT/CR), the California Verbal Learning Test - clusters (CVLT/CL), and the Brown-Peterson Consonant Trigram Auditory Memory Task (CCC) were capable of differentiating responding (rTMD) versus non-responding TMD (nrTMD) patients prior to treatment even after controlling for eleven psychosocial and clinical variables used in the analysis and other four included in the design. Among the best predictors, the CCC in a backward stepwise logistic regression was chosen as the best predictor of all neuropsychological tests. Three psychosocial variables were also found to be good
predictors: a) sleep, b) depression, c) fatigue, and d) income. Among clinical variables, only pain on chewing pre-treatment was found to be predictors. Our secondary objective was to assess if the neuropsychological and psychosocial profile between irritable bowel syndrome (IBS) and nrTMD patients were much more similar than those between the IBS and the rTMD group. On average, both nrTMD and IBS patients did worse in the neuropsychological and psychosocial tests than the rTMD with higher memory deficits, higher levels of depression, sleep disorders and fatigue, and lower energy levels. Taken in combination, the data suggested that neuropsychological differences exist between the nrTMD versus rTMD patients. In addition, it was also demonstrated that two unrelated chronic pain conditions (i.e. IBS and nrTMD) may share similar neuropsychological characteristics. Therefore, solid and reproducible evidence was provided with the use of neuropsychological tests in favor of the biopsychosocial model of chronic pain. This study also supported the multidisciplinary management approach for patients with temporomandibular disorders.

KEYWORDS: temporomandibular disorders, temporomandibular joint, chronic pain, pain, irritable bowel syndrome, neuropsychology, psychology, psychometric tests, sleep disorders, depression, comorbidity, treatment
I would like to thank the my advisory and thesis committee for all the support and guidance throughout this work: Prof. Howard Tenenbaum, Prof. David Locker, Prof. Harvey Moldofsky, Prof. James Fricton, Prof. David Mock, Prof. Adele Csima, Prof. Thuan Dao, and Prof. Brian Levine. I would also like to extend my gratitude to Dr. D. Stuss, Dr. Brenda Toner, Martha Clarke, Lori Mockler as well as all dentists and staff of the Mount Sinai Hospital Orofacial Pain Clinic and the Faculty of Dentistry TMJ Clinic.

I would like to dedicate my work to my beloved wife Patricia and my beautiful son Gabriel. To my parents and family members, Jose, Ercilia, Eunice e Paulo, for their eternal love and support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>xi</td>
</tr>
</tbody>
</table>

## I. REVIEW OF THE LITERATURE

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Introduction</td>
<td>1</td>
</tr>
<tr>
<td>b) Rationale for neuropsychological testing in TMD patients</td>
<td>8</td>
</tr>
<tr>
<td>i) Definition</td>
<td>8</td>
</tr>
<tr>
<td>ii) Clinical Relevance</td>
<td>8</td>
</tr>
<tr>
<td>iii) Summary &amp; critical appraisal</td>
<td>11</td>
</tr>
<tr>
<td>c) Rationale for psychosocial assessment</td>
<td>13</td>
</tr>
<tr>
<td>i) Definition</td>
<td>13</td>
</tr>
<tr>
<td>ii) Clinical Relevance</td>
<td>13</td>
</tr>
<tr>
<td>iii) Assessment of psychosocial variables</td>
<td>17</td>
</tr>
<tr>
<td>iv) Summary</td>
<td>18</td>
</tr>
<tr>
<td>d) Rationale for Sleep Disorders Assessment</td>
<td>20</td>
</tr>
<tr>
<td>i) Definition</td>
<td>20</td>
</tr>
<tr>
<td>ii) Clinical Relevance</td>
<td>20</td>
</tr>
<tr>
<td>iii) Summary</td>
<td>23</td>
</tr>
<tr>
<td>e) Rationale for Irritable Bowel Syndrome</td>
<td>24</td>
</tr>
<tr>
<td>i) Definition</td>
<td>24</td>
</tr>
<tr>
<td>ii) Clinical Relevance</td>
<td>25</td>
</tr>
<tr>
<td>iii) Summary</td>
<td>27</td>
</tr>
<tr>
<td>f) Treatment outcome studies on TMD</td>
<td>28</td>
</tr>
<tr>
<td>i) Clinical Relevance</td>
<td>28</td>
</tr>
<tr>
<td>ii) Psychosocial variables as predictors</td>
<td>30</td>
</tr>
<tr>
<td>iii) Neuropsychological variables as predictors</td>
<td>37</td>
</tr>
<tr>
<td>iv) Summary</td>
<td>39</td>
</tr>
<tr>
<td>g) Rationale for pain assessment as the main outcome measure</td>
<td>42</td>
</tr>
<tr>
<td>i) Definition and importance</td>
<td>42</td>
</tr>
<tr>
<td>ii) Pain assessment</td>
<td>42</td>
</tr>
<tr>
<td>iii) Subjective pain reports and word descriptors</td>
<td>44</td>
</tr>
<tr>
<td>iv) Memory for pain</td>
<td>54</td>
</tr>
<tr>
<td>v) Pain assessment in TMD</td>
<td>59</td>
</tr>
<tr>
<td>vi) Other treatment outcome measures used in patients with chronic pain</td>
<td>60</td>
</tr>
</tbody>
</table>
II. STATEMENT OF THE PROBLEM ................................................................. 68
   a) Problem 1 ......................................................................................... 68
   b) Problem 2 ......................................................................................... 70

III. OBJECTIVES ....................................................................................... 71

IV. HYPOTHESES ....................................................................................... 72

V. RESEARCH DESIGN AND METHODS ....................................................... 73
   a) Experimental design ........................................................................ 73
   b) Treatment outcome measures ......................................................... 73
   c) Improvement criteria ........................................................................ 76
   d) Description of the populations studied ............................................ 77
      i) Temporomandibular disorder group (Groups I and II):
         inclusion criteria ........................................................................ 77
      ii) TMD exclusion criteria (Groups I and II): .................................. 78
      iii) Irritable bowel syndrome group (Group III): inclusion criteria .... 79
      iv) Irritable bowel syndrome (Group III): exclusion criteria .......... 80
      v) Non-pain population (Group IV): rationale .................................. 80
      vi) Non-pain population (Group IV): inclusion criteria ................. 81
      vii) Non-pain population (Group IV): exclusion criteria ............... 81
   e) Instruments ....................................................................................... 82
      i) Neuropsychological assessment ................................................... 83
      ii) Psychosocial assessment .............................................................. 90
      iii) Clinical examination ................................................................ 95
   f) Confounders ..................................................................................... 104
   g) Data analysis .................................................................................... 105
   h) Sample size calculation .................................................................. 107
   i) Examiner .......................................................................................... 108
   j) Informed consent, confidentiality and budget justification .............. 109
   k) Pilot testing ...................................................................................... 110
   l) Protocol for the TMD group ............................................................. 110
      i) Patient selection .......................................................................... 111
      ii) Wash-out period and scheduling of appointments ..................... 112
      iii) Neuropsychological and psychosocial assessment .................... 112
      iv) Clinical examination & treatment regimen .............................. 118
      v) Follow-up questionnaire ............................................................. 119
   m) Protocol for IBS patients ................................................................ 120
   n) Protocol for non-pain patients ......................................................... 121
VI. RESULTS

122

a) Descriptive data on recruitment and follow-up rates .............................................122
b) Assessment of missing values ..............................................................................123
c) Sociodemographic description of the populations studied at baseline ..........124
d) Verbal versus non-verbal improvement criteria .....................................................125
e) Determining the best TMD treatment outcome predictors (Hypothesis I) .....126
   i) Neuropsychological tests ......................................................................................127
   ii) Psychosocial tests ...............................................................................................129
   iii) Assessing the role of confounders .................................................................131
   iv) Clinical signs and symptoms: TMJ and masticatory muscle palpation ..........136
   v) Intra-oral examination and pain intensity ..........................................................139
   vi) Assessing TMD predictors controlling for confounders ..................................142
   vii) Odds ratio between TMD treatment outcome predictors controlling for confounders ........................................146
   viii) Selecting TMD treatment outcome predictors using logistic regression ........150
f) Associations between the rTMD and nrTMD versus the IBS groups ..........153
   i) Neuropsychological and psychosocial tests: nrTMD versus IBS patients ......153
   ii) Confounders: nrTMD versus IBS ......................................................................156
   iii) Neuropsychological and psychosocial tests: rTMD versus IBS .................160
   iv) Confounders: rTMD versus IBS ........................................................................163

VII. DISCUSSION

171

a) Sample size, primary diagnosis, recruitment rate and follow-up .........................171
b) Sociodemographic description of the populations studied at baseline .................172
c) Pain improvement ..................................................................................................174
d) Predictors of treatment outcome for TMD patients ............................................178
   i) Neuropsychological tests ....................................................................................179
   ii) Psychosocial variables .......................................................................................184
   iii) Confounders ......................................................................................................191
   iv) Clinical examination variables as predictors of treatment outcome ..................194
   v) Assessing neuropsychological predictors controlling for confounders ..........201
   vi) Selection of the best predictors among the predictors of TMD treatment outcome ........................................207
e) Associations between the responding and non-responding
   TMD vs. the IBS groups ................................................................. 208
   i) Social and demographic distribution variables .......................... 209
   ii) Neuropsychological and psychosocial tests ............................. 210
f) The non-pain group (Group IV) versus TMD and IBS patients ........ 213
g) Sources of biases, clinical implications and future considerations .... 216

VIII. CONCLUSIONS .............................................................................. 225

X. BIBLIOGRAPHY .................................................................................. 237
LIST OF TABLES

TABLE I: SOCIODEMOGRAPHIC DESCRIPTION OF THE POPULATIONS STUDIED AT BASELINE ............................................................... 125

TABLE II: ASSOCIATIONS BETWEEN VERBAL AND NON-VERBAL PAIN IMPROVEMENT.....126

TABLE III: NEUROPSYCHOLOGICAL AND PSYCHOSOCIAL TEST RESULTS BETWEEN RESPONDING (GROUP I) AND NON-RESPONDING TMD (GROUP II) PATIENTS ..........130

TABLE IV: CONFOUNDERS, RESULTS BETWEEN RESPONDING (GROUP I) AND NON-RESPONDING TMD (GROUP II) PATIENTS .................................................................................135

TABLE V: TMJ AND MASTICATORY MUSCLE PALPATION RESULTS BETWEEN RESPONDING (GROUP I) AND NON-RESPONDING TMD (GROUP II) PATIENTS ...............................138

TABLE VI: INTRA-ORAL EXAMINATION AND PAIN INTENSITY RESULTS BETWEEN RESPONDING (GROUP I) AND NON-RESPONDING TMD (GROUP II) PATIENTS ..........141

TABLE VII: ODDS RATIO AND VARIATION OF ODDS RATIO OF NEUROPSYCHOLOGICAL TESTS VERSUS CONFOUNDERS: RTMD (GROUP I) VERSUS NRTMD (GROUP II) ......145

TABLE VIII: ASSOCIATIONS BETWEEN TREATMENT OUTCOME PREDICTORS FOR TMD: NEUROPSYCHOLOGICAL (COGNITIVE) TESTS .................................................................149

TABLE IX: NEUROPSYCHOLOGICAL TEST RESULTS BETWEEN NON-RESPONDING (GROUP II) AND IRRITABLE BOWEL SYNDROME (GROUP III) PATIENTS ..............155

TABLE X: CONFOUNDERS, RESULTS BETWEEN NON-RESPONDING TMD (GROUP II) AND IBS PATIENTS (GROUP III) ..............................................................157

TABLE XI: ODDS RATIO AND VARIATION OF ODDS RATIO* OF NEUROPSYCHOLOGICAL TESTS VERSUS CONFOUNDERS: NRTMD (GROUP II) VERSUS IBS (GROUP III) ......159

TABLE XII: NEUROPSYCHOLOGICAL TEST RESULTS BETWEEN RESPONDING (GROUP I) AND IRRITABLE BOWEL SYNDROME (GROUP III) PATIENTS .........................162

TABLE XIII: CONFOUNDERS, RESULTS BETWEEN RESPONDING TMD (GROUP I) AND IBS PATIENTS (GROUP III) ..............................................................164

TABLE XIV: ODDS RATIO AND VARIATION OF ODDS RATIO OF NEUROPSYCHOLOGICAL TESTS VERSUS CONFOUNDERS: RTMD (GROUP I) VERSUS IBS (GROUP III) ......166

TABLE XV: NEUROPSYCHOLOGICAL VARIABLES: COMPARISON BETWEEN AND AMONG GROUPS I, II, III, IV .................................................................................................170
LIST OF FIGURES

FIGURE 1: DIAGRAM OF THE RESEARCH PROTOCOL ..................................................111

FIGURE 2: RESULTS OF LOGISTIC REGRESSION ANALYSIS: BACKWARD STEP PREDICTORS
OF TMD TREATMENT OUTCOME (NEUROPSYCHOLOGICAL-COGNITIVE TESTS,
BEHAVIORAL VARIABLES, SLEEP AND CONFOUNDERS) ........................................152
LIST OF APPENDICES

IX. APPENDICES ................................................................................................................................... 228

APPENDIX I: RECRUITMENT LOG OF ALL NEW OROFACIAL PAIN PATIENTS ............................. 228
APPENDIX II: NUMBER AND PROPORTION OF MISSING VALUES PER VARIABLE ...................... 229
APPENDIX III: ASSOCIATIONS BETWEEN VARIABLE WITH MISSING VALUES
EXCEEDING 5% OF VALID CASES VERSUS RESPONDING (GROUP I) AND
NON-RESPONDING TMD (GROUP II) .................................................................................................... 230
APPENDIX IV: CORRELATIONS BETWEEN MULTIPLE CONFOUNDING VARIABLES
IN THE CVLTCR (GROUP I VERSUS II) ................................................................................................ 231
APPENDIX V: GLOSSARY OF ABBREVIATIONS USED IN THIS STUDY ...................................... 232
I. REVIEW OF THE LITERATURE

a) Introduction

Temporomandibular disorders is a collective term embracing a number of chronic pain problems that involve the masticatory musculature, the temporomandibular joint, or both. TMD has been identified as a major cause of nondental pain in the orofacial region and is considered a subclassification of musculoskeletal disorders (Bell, 1989). In the past, TMD signs and symptoms were thought to be related to occlusal irregularities. This belief resulted in diagnostic errors, unnecessary irreversible dental treatments, and treatment failures. Based on both clinical and basic research, new concepts of pain and a broader understanding of chronic pain have resulted in the use of reversible treatments and multidisciplinary teams for pain management. In 1982, the American Dental Association (ADA) held a conference on TMD and stressed the need for a better classification system as well as for better epidemiologic, diagnostic, and treatment data (Laskin et al., 1983). Because of the social and economic impact of TMD, there has been an increasing recognition of the need for epidemiologic studies of the prevalence and consequences of chronic and recurrent pain in human populations. Indeed, epidemiology of orofacial pain is one of the National Institutes of Dental Research priorities (Ogden, 1991).
The current cross-sectional epidemiologic studies on TMD in specific populations show that 75% have at least one sign of joint dysfunction (joint noise, tenderness, etc.) and about 33% have at least one symptom (face pain, joint pain, etc) (Rugh and Solberg, 1985; Schiffman and Fricton, 1988). However, the results vary considerably from study to study because of differences in data collection and inclusion of signs and symptoms, such as headache, and neck pain, that may or may not be related to TMD. In a study using a telephone administered symptom questionnaire in 148 Canadian adults, Locker & Slade, 1989, observed that in 63.5% of the subjects, symptoms were reported; while in 88.1% signs were found.

Although small sex differences have been found in some epidemiologic studies, recent clinical tabulations report a female-to-male ratio of 3:1 to 9:1 in persons seeking care for TMD (McNeill, 1985; Centore et al., 1989). TMD is often self-limiting and fluctuate over time as suggested by the declining incidence with age. Unfortunately, knowledge regarding the natural history or course of TMD is limited (Rasmussen, 1981; Nickerson and Boering, 1989). Finally, signs and symptoms of TMD generally increase in frequency and severity from the second through the fourth decade of life (Laskin et al., 1983).

Few epidemiologic studies used samples drawn from the general population. In a Canadian study (Locker & Slade, 1988), 1,002 adult patients were contacted by the use of a random digit dialing method. According to the results reported, 68% completed a symptom questionnaire; 49% answered positively to at least 1 of the 9 questions about symptoms. Functional pain or pain while resting was reported by 13%. Women and
people in the younger age groups were slightly more likely than men and older age groups to have one or more symptoms. The reporting of symptoms was significantly associated with potential risk factors, such as frequent stress. According to the case definition utilized in this study, 3.5% to 9.7% needed treatment.

The current treatment modalities for chronic TMD are guided by a biomedical model of disease. The biomedical model assumes that an individual’s complaints and ailments result from a specific disease state represented by disordered biological processes. Management is directed toward correcting organ dysfunction or disease. Many treatments for TMD based on this model have been proposed: occlusal adjustment, occlusal splints, orthodontics, and TMJ surgery. The biomedical model has been criticized for its failure to account for psychological and psychosocial variables in health and disease and their dynamic interaction with pathophysiologic factors. Specifically, problems arise when the patient’s complaints and ailments are not commensurate with the degree of observable signs of the disease. In contrast to the biomedical model, the biopsychosocial model focuses on illness, the result of a complex interaction of biologic, psychological, and social variables. The biopsychosocial concept of illness permeates behavioral medicine and has influenced research that examines the conditions under which some individuals become sick, while others do not; and why some patients respond to a variety of treatments for the same condition, while others do not benefit appreciably from any. In the biopsychosocial model, the term “treatment” is replaced by “management”, considering that the cause of the condition is not known. The
management of the condition is directed towards reduction in symptomatology and restoration of function (Dworkin & Massoth, 1994).

Temporomandibular disorders have unknown etiology with a fluctuating chronic pattern of pain without specific association with the signs or symptoms of the disease (Laskin et al., 1983). TMD can be better understood from an illness perspective in which emphasis is placed on the integration of biologic, psychological, and social and/or cultural influences on the experience of chronic pain (Von Korff et al, 1988; Rudy et al, 1995).

Data have been accumulating that TMD is a chronic pain condition that shares many features with other common chronic pain conditions (e.g. headache and back pain). Data from a large sample of headache (N=779), back pain (N=1213), and TMD (N=397) patients show that these conditions are similar regarding clinical pain parameters of intensity, chronicity, frequency, and associated pain-related disability (Von Korff et al, 1992). In addition, affective and behavioral variables indicate that TMD is associated with similar degree of depression and impact on psychosocial functioning as the other two chronic pain conditions. Other common chronic pain conditions, such as irritable bowel syndrome, share with TMD the fact that both are more prevalent among women, have decreased prevalence with age, and are probably self-limiting (Gerke et al, 1988).

The concept of chronic pain dysfunction has been introduced to explain that in chronic pain conditions, the degree of pain-related disability is often not commensurate with the degree of observable disease. Patients may have impairment and disability due to a disease-related disruption of an organ or system; however, in TMD, there may be a
psychosocial dysfunction, or disruption in functioning at the interpersonal or social level. The assessment of psychosocial functioning in TMD has received increased attention. The Multidimensional Pain Inventory (MPI) developed by Turk and colleagues (Rudy et al, 1995) is a widely used pain measure, which distinguishes between psychosocially “dysfunctional” chronic pain patients and their functional counterparts, termed “adaptive copers.” Although TMD patients are not significantly different along physical parameters, dysfunctional patients show significantly elevated depression and report significantly more physical symptoms than “adaptive copers.” In addition, both depression and somatization have been heavily implicated in chronic pain, including TMD (Dworkin et al, 1990d; McCreary et al, 1992).

Despite the high success rate, between 70 to 98%, it is also clear that 2 to 30% of TMD patients do not improve and may in fact be non-responding to the management approach (Greene & Laskin, 1983). This failure has usually been attributed to the combination of behavioral, psychological and psychosocial factors associated with TMD and their interaction with pathophysiologic factors (Fricton et al., 1988).

To date, little has been learned about those patients who are non-responding (Fricton & Olsen, 1996; Gerschman et al., 1987), since most investigations have, understandably, focused on developing successful management, or to understanding the TMD population as a whole (Greene & Laskin, 1983). Thus, it is clear that there is a pressing need to develop an understanding of those individuals who do not improve since it would appear that those are the patients who pose the greatest challenge to therapy.
In relation to the above, a recent investigation carried out in this Research Unit has demonstrated that patients suffering from post-traumatic TMD (pTMD) as a result of a Motor Vehicle Accident (MVA) have significantly increased prevalence and magnitude of cognitive and neuropsychological deficits (i.e. memory, attention, reaction time deficits) in comparison to patients with non-traumatic or idiopathic TMD (iTMD)(Goldberg et al., 1996). Importantly, it has also been shown that successful management in patients who develop pTMD is significantly reduced as compared to the iTMD patients (48% vs. 80% respectively)(Romanelli, Mock & Tenenbaum, 1992). Accordingly, we suggest that neuropsychological deficits may either play an integral role in mediating poor treatment outcome, or at least may be predictors of this in the post-traumatic population. By extension, we predict that iTMD patients with poor treatment outcome may also have some degree of cognitive or neuropsychological impairment in comparison with those who do recover, irrespective of management rendered.

Another interesting finding was that the scores in the pTMD population were comparable to those of patients with closed head injury (Stuss et al., 1989a, 1989b) at least suggesting that these two populations might have similar etiologies. This finding is very significant, because it provides mechanisms to differentiate among different TMD populations and as an aid for diagnostic classification. Finally, it also helps in the development of appropriate multidisciplinary managements for different TMD groups. Therefore, there is the need to increase the number of studies in dentistry dealing with neuropsychological tests.
Finally, it will be also interesting to assess common neuropsychological characteristics between two chronic pain conditions of different origins, in the case of this investigation: temporomandibular disorders and irritable bowel syndrome, in order to support or not the concept of the biopsychosocial model for temporomandibular disorders, particularly in the non-responding TMD group.
b) Rationale for neuropsychological testing in TMD patients

i) Definition

Neuropsychology is generally understood to be the study of the relation between brain function and behavior. Neuropsychological assessment has traditionally focused on determining specific changes in mental processes in patients with discrete lesions following normal development. In this manner, neuropsychology in clinical practice has enabled practitioners to determine the location of the insult or the disease as well as the functional capacities of patients in treatment. Neuroscientists have benefited from human neuropsychological data, since studies of the relation between brain lesions and specific patterns of functional deficit have made substantial contributions to the understanding of the role of specific neuroanatomic regions in normal mental processes (Kolb & Whishaw, 1989).

ii) Clinical Relevance

There are many applications of neuropsychology in psychiatry, including the identification of brain lesions in psychiatric patients, the evaluation of cognitive
deterioration over time, and the advancement of theory regarding the neuroanatomic localization of the symptoms of various neuro-psychiatric disorders. Neuropsychological test batteries, when constructed properly, assess a broad range of functions, including among others, perception in all sensory modalities: attention, learning and memory, motor skills, verbal and nonverbal skills, comprehension and expression of language, spatial abilities, laterality, abstraction, and "executive" functions. One of the most important contributions of neuropsychological assessment is that it makes possible an objective evaluation of behavior in the context of the ability to perform basic tasks. In a study of the neuropsychological profile of patients with good recovery after closed head injury, Stuss et al. (1985, 1989a) demonstrated that those patients show consistency of performance, which shows reproducibility of the results. When applied properly, a battery of neuropsychological tests provides the clinician or investigator with an objective description of what areas of behavior and cognition are likely to be a problem for the psychiatric patient and what areas are not. In this manner, neuropsychological data serve as an indicator into the everyday mental processes of the psychiatric patient. Neuropsychological test batteries are useful in making general conclusions about the presence of impairment in individuals without psychiatric disorders, but they are not flexible enough to allow the investigator to assess the variety of functions that may underlie a performance deficit on a complex cognitive task (Kolb & Whishaw, 1989).

The greatest contribution of the neuropsychological evaluation of patients with neuro-psychiatric disorders (e.g., Alzheimer's disease) may be that it provides important, objective data about the mental deficiencies that shape our patients's lives. While
neuropsychological tests may serve to inform specific neuropathological models of psychiatric disorders by comparing the performance of psychiatric patients and patients with brain lesions, this role is likely to be replaced in the future because of the tremendous advances in imaging technologies. However, images of the structure and regional activation of the brains of our patients will not provide us with information about their difficulties with mental processes or about their cognitive strengths that can be used to facilitate treatment (Kolb & Whishaw, 1989).

The identification of specific cognitive deficits in psychiatric disorders may be a powerful predictor of the course of illness. Patients of a particular diagnostic group who also have global cognitive impairment are likely to have worse outcomes than similarly diagnosed patients who perform normally on the neuropsychological tests. In a group of patients assessed in a psychiatric emergency room, cognitive deficit was the single best predictor of referral for inpatient hospitalization; it was even superior to the patient’s diagnoses (Galankyer & Harvey, 1992). In some disorders, such as schizophrenia and Alzheimer’s disease, cognitive deficits as assessed by a battery of neuropsychological tests may predict the onset of illness. It has been demonstrated that general impairment on tests assessing information, memory, and concentration serves as a better tool for the prediction of the eventual development of Alzheimer’s disease in normal elderly volunteers than prior head injury, age of the mother when the subject was born, smoking, or family history of Alzheimer’s disease (Katzman et al., 1989).

Similar to many major psychiatric disorders, the pattern of cognitive deficits among patients in a single diagnostic group is heterogeneous. Not all schizophrenic
patients perform poorly on tests of verbal memory or the Wisconsin Card Sorting Test, and not all depressed patients perform poorly on tests of psychomotor speed. The identification of stable patterns of deficit on neuropsychological tests within a disorder may contribute to the development of hypotheses about the differing etiologies of the disorder (Shallice et al., 1991; Green 1992).

There are few empirical data on the relation between neuropsychological deficit and response to medication, psychotherapy, and treatment setting. However, baseline cognitive impairment may be an important predictor of treatment outcome. It seems possible that the different responses to treatment among patients with psychotic, affective, and anxiety disorders may be related to their pretreatment level of cognitive functioning, and that treatment regimens which are suited specifically to an individual's pattern of cognitive deficits and abilities may be more effective (Spaulding & Sullivan, 1991). An example of an improved psychotherapeutic strategy based on data from cognitive studies can be found in the treatment of severe anorexia nervosa. Cognitive behavioral therapy has been more successful than traditional approaches (Garner et al., 1983).

iii) Summary & critical appraisal

In summary, the differential pattern of neuropsychological performance (e.g. attention, learning and memory, motor skills, verbal and nonverbal skills, comprehension and expression of language, spatial abilities, laterality, abstraction, and “executive”
functions) can be used in psychiatric and non-psychiatric populations for objective
evaluation of behavior in the context of the ability to perform basic tasks with good
reproducibility (Stuss et al., 1985; Kolb & Whishaw, 1989). In addition, they can also be
used to assess the need of cognitive rehabilitation, to predict the course of psychiatric and
non-psychiatric illnesses, to reduce the diagnostic heterogeneity within disorders and thus
improve diagnostic classifications, to develop hypotheses about the differing etiologies,
to serve as an aid in the development of treatment options, and to create individualized
treatment approaches (Katzman et al., 1989; Spaulding & Sullivan, 1991; Galankyer &
Harvey, 1992).
c) Rationale for psychosocial assessment

i) Definition

Psychosocial assessment is a combination of psychology and social factors which seems to be inseparable and might either cause, precipitate or aggravate chronic pain (Schwartz et al., 1979). Psychology is the science which deals with the mind and with mental and emotional processes. Social processes are related to human beings and their living together (Webster's New World Dictionary, 1989).

ii) Clinical Relevance

The rationale for psychological assessments of patients with TMD is based upon the assumption that psychological factors may predispose, serve as etiologic factors, and maintain or be a consequence of TMD. Evidence has been accumulating since the 1950's that psychological factors are of concern in certain subgroups of patients with TMD (Greene et al., 1982).

In addition, data from Dworkin (1994) indicate that TMD shares the major characteristics of other common chronic pain conditions, particularly headache and back pain. Clinically significant anxiety has been reported in 17% and 26% of patients with
TMD (Fricton et al., 1985; Gerschman et al., 1987); while depression was found in 18% (Gerschman et al., 1987). On the other hand, depression is only found in non-pain population in about 6% (Nielsen and Williams, 1980). Whenever these conditions are significantly present, they should be recognized and properly managed. Recently, more studies have been done comparing different psychological characteristics among subgroups of TMD. It was found that psychological factors are more frequent in patients with muscle-related disorders than in those with joint-related disorders (Eversole and Machado 1985; Butterworth & Deardorff, 1987; McCreary et al., 1991).

Greene et al., 1982, reviewing the role of psychological factors in TMD, stressed the importance of multiple factors in the etiology of the disease, including personality traits, anxiety, depression, stress-induced muscle hyperactivity, illness-behavior, and doctor-patient relationship. The multifactorial etiology of TMD and the complete examination of all physical, emotional, and behavioral factors involved in the disease were also emphasized by Rugh and Solberg, 1976.

Several investigators have described depression as an etiological factor in Myofascial Pain-Dysfunction Syndrome (MPD). They report that some of the complaints of MPD patients are weight loss, difficulty in eating, difficulty in sleeping, and a reduction in social activity. But these symptoms may also be the result of chronic pain. Treatments for depression alleviate pain, but a number of treatment modalities also have the same effect. In addition, similar depression prevalence between TMD and other chronic pain syndromes is not a confirmation of etiology, because the chronic pains may be the initiating factor, and depression one of the consequences.
In addition, they can also be used as evidence to support the biopsychosocial model for non-responding TMD patients and for the development of better classification systems. Several diagnostic criteria have been developed to classify patients with temporomandibular disorders (TMD) (Dworkin & LeResche, 1992). These systems typically diagnose TMD patients on the basis of the presence or absence of a set of somatic symptoms (e.g., pain and joint noises) and physical signs (e.g., joint sounds, limitations of jaw opening, tenderness of muscles to palpation and radiographic findings). Neither the reliability of the procedures used to evaluate patients nor the validity or clinical utility of any of these diagnostic systems have been demonstrated (Dworkin et al, 1988, 1990a).

Psychological inventories like the Minnesota Multiphasic Personality Inventory (MMPI) and Symptom Checklist-90 revised (SCL-90R) have been used to identify subgroups of TMD patients (Butterworth and Deardorf 1987; McCreary et al., 1991). Turk (1990) has suggested that psychosocial and behavioral factors (e.g., spouse support, feelings of control, activity levels) play an important role in chronic pain. Using empirically based classification methods, Rudy et al. (1995) subdivided TMD patients into 3 subgroups based on psychosocial and behavioral factors: 'Dysfunctional' (DYS), 'Interpersonally Distressed' (ID), and 'Adaptive Copers' (AC). The first group, the DYS group, was characterized by higher levels of pain, interference of symptoms with life activities, affective distress, and lower levels of activity and feeling of life control. The second one, the ID group, similar to the DYS group, was characterized by little support from significant people in their environment and received a number of negative responses
or assistance to their pain problem by others. The third one, the AC group, was described as having little psychological distress or life interference and control of their lives despite the presence of pain. These groups did not differ on the physical findings associated with TMD (e.g., disc displacements, degenerative conditions of the TMJ). Therefore, treatments that take into account psychosocial and behavioral factors may lead to better treatment outcomes. However, up until now, the clinical utility of psychological classifications have been demonstrated only by few studies (Turk 1990; Dworkin 1992). One way to validate the clinical utility of any classification system is to demonstrate that patients with distinct pre-treatment characteristics will respond differentially on different outcome measures (Fricton & Olsen, 1996; Lipton & Marbach, 1984).

Relationships among TMD patients and biopsychosocial factors, which are relevant to the characterization of TMD as a chronic pain condition, were examined by use of data from longitudinal studies of TMD patients and controls (Dworkin et al, 1989; Dworkin et al, 1991). Dysfunctional TMD patients score significantly higher, in the top 15% to 25% of scores on measures of depression and somatization. However, dysfunctional TMD patients are indistinguishable from functional TMD patients in unassisted vertical range of jaw motion or the total number of joint sounds detected on both vertical and lateral excursions of the jaw. These data indicate that TMD pain dysfunction is not significantly associated with TMD signs commonly used in clinical assessment.
iii) Assessment of psychosocial variables

Nevertheless, the dentist's lack of ability in identifying psychopathological disorders have obscured the importance of its assessment (Oakley et al., 1989). Solberg 1986, evaluating the difficulty in managing TMD, stated that "the dentists uneasiness in coping with psychosocial factors associated with multifactorial disease has prevented the correct evaluation and management of them." In addition, the pragmatics of dental practice has hindered a thorough understanding of the whole patient. The lack of a correct diagnosis and the consequent treatment failure in TMD lead to an exacerbation of the chronic pain cycle.

It has already been shown that physicians misdiagnose stressful life events in 76% of patients and psychiatric disturbances in 34%, and dentists fail or tend to overdiagnose psychological problems during the clinical history alone (e.g. anxiety, life stress, and depression) (Fricton et al., 1988). Therefore, in order to overcome the deficiencies of clinical history, many instruments have been developed for assessing psychological and behavioral conditions. Despite the lack of a recognized gold standard, the DSM-IV is the most commonly used gold standard for diagnosis of psychological and behavioral conditions in medical disorders. However, it does not adequately address certain aspects related to chronic pain such as illness behaviors, irrational beliefs about pain and oral habits and is influenced by the ability and expertise of the clinician (Fricton et al., 1988). The Minnesota Multiphasic Personality Inventory (MMPI) is the most widely studied psychodiagnostic inventory. Other simpler instruments, such as the Beck Depression
Inventory and the Zung Self-Rating Depression scale, have also been used. When comparing sensitivity and specificity of these questionnaires using the DSM-IV as the gold standard, it was found that the simpler questionnaires were superior to the MMPI, probably due to the fact that simpler instruments increase patient compliance (Turner & Romano, 1984). Specifically designed for TMD patients, the TMJ scale is a 97-item questionnaire which assesses factors believed to be relevant for TMD (e.g. psychological factors and stress). The sensitivity and specificity of this instrument is comparable to other screening instruments; however, it was tested against the Symptom Checklist (SCL-90) and the Derogatis Stress Profile instead of the DSM-IV (Levitt, 1990). Other specific instruments for TMD psychological and behavioral assessment are the IMPATH:TMJ (Fricton et al., 1987) and the RDC/TMD Axis II.

iv) Summary

A number of studies have reported that in addition to any discrimination of patients based on somatic factors, TMD patients may differ on important psychological characteristics (Greene et al., 1982; Dworkin et al., 1989, 1991; Rudy et al, 1995). These studies are difficult to compare due to a lack of universally accepted population definition, as well as defined exclusion/inclusion criteria, and different methodologies employed. However, the majority of them have suggested that psychological factors may play a significant role in TMD and have identified subgroups of patients based on their responses to psychological assessments; however, few have reported on the clinical
outcome of these classifications (Dworkin & LeResche, 1992; Fricton & Olsen, 1996). Rudy et al. (1995) suggested that TMD patients may differ not only on objective measures, but also on psychological characteristics. In other words, subgroups of TMD patients may differ on psychosocial and behavioral features. This may offer an explanation why some patients improve with treatment while others do not.

Despite the fact that a significant number of studies have been done assessing psychological and behavioral problems (e.g. back pain), few have been used for evaluating subgroups of TMD (Fricton & Schiffman, 1987; Dworkin & LeResche, 1992; Fricton & Olsen, 1996). In addition, in a review article of behavioral assessment of chronic orofacial pain (Keefe & Beckham, 1990), none of the studies reviewed specified the validity of those measurements. Therefore, longitudinal studies are needed in order to assess the validity of those measurement instruments. In addition, despite our understanding of the role of psychosocial factors in the etiology of temporomandibular pain and pain in general, little is yet known about the factors contributing to the perpetuation of chronic pain. Previous research has not emphasized the importance of psychosocial factors in predicting outcome for patients with chronic pain. Finally, there is still the need for longitudinal studies assessing particularly the role of psychosocial variables as perpetuating factors, rather than etiologic.
d) Rationale for Sleep Disorders Assessment

i) Definition

Sleep disturbances as a behavior can be dangerous and destructive, especially in patients with chronic pain and disturbed sleep. Sleep by definition is divided into two phases: the rapid eye movement phase (REM) and the non-rapid eye movement phase (non-REM). The non-REM phase is divided into four stages with each numerically higher stage representing a deeper stage of sleep. Stage 1 is considered a light sleep, stage 2 is intermediate, and stages 3 and 4 are deep sleep. REM sleep is characterized by rapid eye movements, muscle twitching, changing pulse, changing respiratory rates, and the phenomenon of dreaming (Sheldon, 1996).

ii) Clinical Relevance

Apparently, REM sleep plays an important role in learning and short- and long-term memory (Karni, 1994; Wilson & McNaughton, 1994). Bruxism occurs primarily in stage 2 non-REM and REM. Time spent in stage 2 non-REM and REM is proportionately higher with stress and sleep disturbance concomitant with loss of restful restorative sleep, myalgia and headaches (Austing, 1997).
Obstructive sleep apnea is a potentially life threatening sleep disturbance. Obstructive sleep apnea is characterized by snoring, headaches, multiple episodes of discontinue respiration (apnea), and excessive sleepiness during the day (Dexter, 1984). The excessive sleepiness has been blamed for such disasters as the Exxon Valdez oil spill, the Three Mile Island nuclear reactor accident, Chernobyl nuclear reactor accident, and thousands of work related and motor vehicle accidents every year (Blau, 1990). Blood oxygen saturation can be lowered with each apnea episode, sometimes below 50%, and this can place incredible stress on the heart and the oxygen-sensitive brain (Austing, 1997).

The relationship between disturbed sleep and pain, fatigue and emotional distress has been revealed by systematic clinical observations. Fibromyalgia patients commonly report that light unrefreshing sleep is followed by generalized aching or stiffness, and profound exhaustion. The symptoms of myalgia, fatigue, unrefreshing sleep and emotional distress are associated with various syndromes, such as irritable bowel syndrome, chronic fatigue, and myofascial pain syndrome. Recent studies of patients with fibromyalgia demonstrate that sleep is intimately related to somatic symptoms that muscle aching and tender points were the variables most strongly associated with non-restorative sleep (Moldofsky, 1993a). These studies show that the amalgam of disordered sleep physiology, chronic fatigue, diffuse myalgia, and cognitive and behavioral symptoms constitute a “non-restorative syndrome.” The chronic fatigue syndrome and fibromyalgia have similar disordered sleep physiology, namely an alpha rhythm disturbance (7.5-11 Hz) in the electroencephalogram (EEG) within non-rapid eye
movement (NREM) sleep that accompanies increased nocturnal vigilance and light, unrefreshing sleep (Moldofsky, 1993b). In a multicenter study, fibromyalgia patients more commonly reported awakening tired (i.e. nonrefreshed sleep), widespread pain, fatigue, morning stiffness, paresthesias, anxiety, headache, and irritable bowel syndrome than did the controls with various rheumatic disorders (Wolfe et al, 1990).

Another important aspect that has not been evaluated is the relationship between sleep disorders and TMD and their prognostic utility. The accumulated clinical and sleep laboratory studies indicate that sleep is intimately associated with musculoskeletal pain, fatigue and psychological distress in patients with fibromyalgia. Accordingly, modification of any of these parameters should have an effect not only on sleep, but also on other features of the syndrome. The treatment literature with regard to modulators of psychological distress (hypnotherapy) is consistent with this prediction (Moldofsky, 1993a). Tricyclic medications have also been used to improve sleep and musculoskeletal pain, especially amitriptyline and cyclobenzaprine (Carette et al, 1994). However, both treatments have been used predominantly for fibromyalgia patients, and their validity in TMD patients is still not clear. If fibromyalgia and TMD patients are actually similar entities (Dao et al., 1996), it is expected that treatments aimed to improve sleep and psychosocial functioning will also be effective in TMD. Other chronic diseases, such as chronic fatigue and irritable bowel syndrome, have also been associated with affective disorders and show similarities with fibromyalgia and TMD (Hudson et al, 1992).

A recent publication tried to assess sleep, daytime symptoms, and cognitive performance in patients with fibromyalgia (FM). The patients (n = 10) were age and sex
matched with a noncompliantive comparison group (n = 9). The authors found that patients with FM spent more time in stage 1 sleep. Patients with fibromyalgia (a chronic muscular pain condition) reported greater sleepiness, more fatigue, more pain, more negative mood, lower speed on cognitive performance of complex tasks (i.e., grammatical reasoning, serial addition/subtraction, and a simulated multi-task office procedure) as compared to a noncompliantive comparison group (Côté & Moldofsky, 1997).

iii) Summary

Most studies on sleep disorders and chronic pain conditions are only available in regard to somatic pain. With regard to orofacial pain, to our knowledge, only one study has addressed the association between sleep disorders and TMD as a predictor of treatment outcome (Fricton & Olsen, 1996). Therefore, additional sleep studies are also necessary not only for the TMD population as a whole but also for specific sub-groups of TMD, i.e. responding versus non-responding TMD, in order to assess their validity. Also, they may provide insights into the mechanisms by which cognitive impairments and chronic pain may occur. In addition, if sleep disorders are shown to be valid as disease predictors and to be comparable between two unrelated chronic pain conditions, they will provide valuable clinical information and solid evidence in favor of the biopsychosocial model.
e) Rationale for Irritable Bowel Syndrome

i) Definition

The functional gastrointestinal (GI) disorders, defined as a "...variable combination of chronic or recurrent gastrointestinal symptoms not explained by structural or biochemical abnormalities," are ever-present in society and in physicians' medical offices. Current estimates indicate high frequency rates in western and eastern societies and at least in western society, they comprise the majority of gastroenterologic practice. Similar to what is observed in TMD patients, the disease definition is symptom-based. The frequencies of functional constipation and diarrhea are 3% and 1.6%, respectively, in the US (Thompson et al., 1992).

Irritable bowel syndrome is a functional bowel disorder (i.e., a functional GI disorder division) in which abdominal pain is associated with defecation or a change in bowel habit, and with features of disorders defecation and with distension (Thompson et al., 1992). Functional gastrointestinal symptoms, like TMD, occur in a high proportion of adult Westerners, and 8 to 19% have symptoms consistent with a diagnosis of IBS (Drossman et al., 1993). Irritable bowel syndrome, similar to TMD, is more prevalent among women, has decreased prevalence with age, and is probably self-limiting (Gerke et al, 1988). In a random sample of Britons, recurrent, intestinal-type abdominal pain was found in 20% of women and 10% of men (Heaton et al., 1992).
ii) Clinical Relevance

A review of the literature reveals that IBS is a chronic pain condition that shares many features with other common chronic pain conditions. Myalgia, fatigue, unrefreshing sleep and emotional distress are associated with various syndromes, such as irritable bowel syndrome, chronic fatigue, and myofascial pain syndrome (Moldofsky, 1993a). Irritable bowel syndrome and TMD have also been associated with affective disorders and show similarities with fibromyalgia and chronic fatigue syndrome (Hudson et al., 1992).

These chronic pain conditions are associated with similar impact on psychosocial functioning and health care costs (Dworkin, 1994). While most persons with IBS do not see physicians, the disorder accounts for 20 to 50% of referrals to gastroenterology clinics (Harvey et al., 1983). From six national surveys, Sandler (1990) concluded that 4.7 million Americans (2.9% of the population) had a diagnosis of the IBS. Female, white, and middle-aged persons were most likely to have such a diagnosis. There were between 2.4 and 3.5 million physician visits annually for the IBS and 2.2 million prescriptions were written. These figures underestimate the prevalence of IBS because they do not include people with these symptoms who do not see physicians. Employing the Rome criteria, 9.4% (14% of females and 7.7% of males) were identified in a survey of 8,250 households in the US. Most of them did not consult with a physician. Females with IBS not only are more common than males, but also are more likely to consult a physician (Drossman et al., 1993).
In addition, affective and behavioral variables (i.e., degree of depression) indicate that these chronic pain conditions are similar to each other. However, it remains unclear to what extent IBS symptoms represent normal perception of abnormal function or abnormal perception of normal function. Stressful life events frequently precede the onset of IBS symptoms, or at least the reporting of them to a physician (Ford et al., 1987). Most studies indicate that mood and personality disturbances, psychiatric disease, and illness behavior are more common in IBS patients seen by a hospital physician or specialist than in other patients and normal subjects (Drossman & Thompson, 1992; Thompson, 1993). A retrospective study of previously gathered data determined that major depression and panic disorder were more common in those with IBS than in others (Walker et al., 1993). Psychological and behavioral treatments seem to be useful in IBS patients. Whorwell et al. (1984) treated IBS patients resistant to conventional treatment and reported that hypnotherapy markedly reduced all symptoms, at least in patients under 50 years old with "typical symptoms." Guthrie et al. (1991) treated resistant patients with seven sessions of dynamic psychotherapy plus relaxation-teaching tape cassettes and found that bloating, pain, and diarrhea improved. It has been shown in the literature that cognitive-behavioral treatment (CBT) helps patients with IBS to increase recognition of the role played by attention allocation, personal appraisal style and illness beliefs in chronic pain and psychosomatic disorders. CBT may be particularly useful for IBS, and possibly other chronic pain disorders, because studies have found in IBS patients a high prevalence of anxiety and depression, a high frequency of assertion difficulties, a high need for social approval and perfectionist attitudes, all of which are ameliorated by CBT (Toner, 1994).
Relaxation/stress therapy has been shown also effective in both IBS and TMD patients, because of its effect in reducing autonomic arousal and anxiety (Turk et al., 1993b; Toner, 1994).

iii) Summary

These studies suggest that the many similarities among these four chronic pain conditions (chronic fatigue syndrome, fibromyalgia, irritable bowel syndrome and temporomandibular disorders) such as disordered sleep physiology, chronic fatigue, diffuse myalgia, and cognitive and behavioral symptoms may constitute a "non-restorative syndrome" (Moldofsky, 1993a). My decision to use IBS as a group to be compared with TMD was based on the fact that it is an anatomically unrelated chronic pain disorder connected to the gastrointestinal system, rather than the musculo-skeletal system, as it is the case in the other three groups (Thompson et al., 1992). In addition, IBS patients are generally non-responding sharing many similarities with TMD non-responding ones. However, this comparison has never been made for TMD sub-groups. If we are able to show similar neuropsychological patterns between IBS and non-responding TMD patients, we will be able to provide solid and reproducible evidence not only in favor of the biopsychosocial model of chronic pain, but also to support the concept that these two unrelated chronic pain conditions may be part of a similar chronic pain syndrome with obvious implications in disease classification and treatment modalities (Moldofsky, 1993a).
f) Treatment outcome studies on TMD

i) Clinical Relevance

A recent tendency in the psychologic literature has been to try to identify psychosocial factors predictive of TMD patient's response to treatment (Fricton & Olsen, 1996; Gerschman et al, 1987; Gale & Funch, 1984; Lipton & Marbach, 1984; Millstein-Prentky & Olson, 1979; Schwartz et al, 1979). Therefore, reliable psychological and behavioral measurements for specific sub-groups of TMD, i.e. non-responding TMD patients, is necessary in order to assess their validity.

More recently, the relationship between TMD and psychosocial variables have been examined. McCreary et al, 1992, using the Minnesota Multiphasic Personality Inventory (MMPI). They found that somatization was a significant factor predicting TMD treatment outcome. The authors concluded that unless clinicians attend to the level of somatization in their patients, successful treatment outcome for TMD is less likely. Wilson et al, 1994, investigated the tendency to report nonspecific physical symptoms as a potential influence on patients’ response style during clinical palpation of masticatory muscles. They found that high somatization scores were associated with three times as many positive pain responses to placebo site palpation on the face and head. Several studies have noted poor ability to predict treatment outcomes, calling attention to nonspecific factors such as appreciable rates of spontaneous cures, motivation, doctor-
patient relationship and psychological status that might influence treatment outcome (Okeson & Hayes, 1986; Gale & Funch, 1984).

A series of studies examined the course of TMD pain in relation to diagnosis, clinical signs, and level of psychosocial function. In one study of 232 patients referred for TMD treatment (Dworkin et al., 1990c), a clinical diagnosis of TMD was assigned according to the published diagnostic criteria of Eversole and Machado (1985). In this sample, the absence or presence of TMD pain one year after baseline was not predicted by diagnostic category at baseline, whether the clinical diagnosis was myofascial pain dysfunction (MPD), internal derangement (ID), or degenerative joint disease (DJD). The presence of pain after one year was predicted by gender (female), by baseline somatization score, pain duration, and pain-related disability. Among pain patients, the pain level after one year was not associated with physical diagnosis, but was significantly related to presence of disability days, heightened somatization, anxiety, and depression. In addition, fluctuation of pain was not associated with fluctuation of jaw opening or joint sounds. Masticatory muscles pain reports were just marginally associated with pain and may be influenced by somatization levels (Dworkin et al., 1989).

Dworkin & Massoth, 1994, used recent longitudinal data to explore the possibility that TMD is a recurrent but non-progressive condition. Approximately 60% of the original clinical sample (N=191) still reported at least some TMD pain at 3-year follow-up. They were also followed up longitudinally to examine how changes in pain grade over 3 years were related to clinical signs and level of psychosocial function. All patients showed gradual improvement in jaw-opening over time, as seen for other clinical
variables (including joint sounds) measured longitudinally. No relationship was observed between any clinical signs and pain-related disorders. They also examined the data to determine whether functional versus dysfunctional TMD patients differ over time in pain report and measures of clinical and psychological functioning. Clinical signs and pain report were not correlated for both groups; however, somatization in functional versus dysfunctional TMD patients revealed a contrasting pattern. Functional TMD patients show a close correspondence between pain measures and psychological measures (such as somatization). However, these relationships were independent of physical measures. Dysfunctional TMD patients show psychological changes that are not consistent with pain-related or physical measures.

ii) Psychosocial variables as predictors

Psychosocial factors have been frequently suggested as important risk factors that may delay recovery in patients with temporomandibular disorders. In a study by Fricton & Olsen (1996), subjects with chronic temporomandibular disorders were studied using IMPATH:TMJ prior to their entering an interdisciplinary treatment program to determine six months after which factors were most predictive of outcome. The primary diagnoses were myofascial pain for 47 (50%) and TMJ disc displacement for the remaining 47 (50%). Regression analysis limited the sample to 47 subjects. In the population studied, 89% were females and 11% males (Mean=36.6 years, SD = 11.7, range 16 to 61 years). Patients were treated by a multidisciplinary team with reversible therapy. Treatment
outcome was determined based on significant decreases in the Craniomandibular Index (post-treatment scores were less than one standard deviation or score less than 0.15) and the Symptom Severity Index (30% reduction) from pre-treatment to post-treatment. Low self-esteem, feeling worried, low energy, and sleep activity were identified as useful predictors of treatment outcome for the criterion group. Each were good correlates of depression. The model correctly predicted treatment outcome for 41 of 47 subjects (87%).

In a study by Schnurr et al. (1991), 178 patients with TMD (Eversole and Machado's criteria, 1985) were given a pressure pain threshold and tolerance task and completed the Basic Personality Inventory, the Illness Behavior Questionnaire, the Multidimensional Health Locus of Control, the Perceived Stress Scale, and the Ways of Coping Checklist. Subjects also answered questions pertaining to TMD symptomatology, including chronicity and severity. The patients were categorized as having myogenic facial pain (n=39) and internal derangement (n = 139). The mean age was 27.4 years (SD = 9.7). After conservative treatment with simple jaw exercises and ultrasound, patients were contacted again at 5 months to complete a follow-up questionnaire package similar to the initial questionnaire battery. Percent reduction in average pain intensity (VAS six-point scale, 50% improvement over baseline according to Blanchard and Andrasik's procedure) and perceived TMD severity (50% improvement over baseline) were used as outcome criteria. The data were analyzed with discriminant function analyses. One hundred patients responded to the follow-up questionnaire (percentage of responders = 56.2%). The response rate does not appear to be a function of treatment success, because the percentage of replying patients who believed that treatment was successful (31.4%
was approximately equal to those who thought that it was not successful (34.8%). Forty three percent of the patients reported were classified as improved and tended to be less inclined to accept responsibility for their problems and were slightly better able to distance themselves from their problems than the less improved groups. Patients who reported more than a 50% reduction in TMD severity (11% of the total) indicated that the condition was not associated with an identifiable onset event and that the condition had become moderately worse between onset and first seeking help. Of the five variable sets entered into the discriminant function analyses to predict changes in pain intensity, the only analysis to obtain significance was that for the Ways of Coping Scale.

In a study by Gerschman et al. (1987), one-hundred and thirty patients with dental phobias and 368 patients with chronic orofacial pain were compared for psychological and social variables. Patients had an average age of 47 years (range 5 to 60) and were assessed using the Diagnostic and Statistical Manual (DSM III), Eysenck Personality Inventory (Form B), the Hamilton Anxiety Scale, the Hamilton Depression Scale. DSM-III found in patients with orofacial pain the following diagnosis: neurotic disorders (52.2%), depression being the predominant disorder, affective disorders (16%), personality disorders (11.5), and normal (8.3%). Using the Hamilton Depression Scale, the patients were classified in the following manner: severely depressed (17.9%), 51.7% moderately depressed, 11.5% mildly depressed, 18.9% not depressed. Pain patients showed a greater burden of psychiatric disorders and were more likely to be older, married, have children, be migrants, be less educated, have poorer jobs and be more financially disadvantaged than phobic patients, Gerschman et al. (1987).
Social and cultural factors have been found to influence the meaning of pain, reaction to pain and communication of pain as well as psychophysiological and autonomic functioning. Lipton & Marbach, 1984, reported that the eight factors included in their discriminant function allowed them to correctly classify 79.3% of temporomandibular subjects (n = 68, 76.5% women) with a mean of age of 40 years (14.3 SD) and 13.1 years (2.8 SD) of education. In order to evaluate the level of psychological distress, Langner 22-item index was used. Sociocultural variables (age, marital status, position in the family, sex, family income, years of education, ethnic group, generation American, perceived ethnicity) as well as social group attitude variables (family tradition, authority, friendship solidarity, ethnic exclusivity) were also analyzed. The treatments employed were all reversible: TMJ injections, local anesthetic nerve blocks, physical therapy (exercises, vapocoolant sprays, and moist heat), non-steroid anti-inflammatory drugs, psychotropic medications (antidepressants, tranquilizers). The treatment outcome measure was the degree of improvement judged by two evaluating clinicians. The reported success rate was 57%. The factors were effect of previous treatment (discriminant analysis coefficient = 0.49), number of doctors previously consulted (0.46), pain duration (-0.53) and number of doctors previously consulted (-0.51). The authors failed to cross-validate the equation on an independent sample, and thus, confidence in their findings is diminished.

In a study by Gale & Funch (1984), 42 chronic TMD pain patients were assessed for predictors of short-term (after treatment) and long-term success (two-years after). The patient population was primarily comprised of women (93%), with a median pain
duration of 3.9 years and a mean pain duration of 6.0 years, and had on average 2.8 previous treatments. The treatment employed were behavioral techniques. The baseline tests were the Wallston’s Health Locus of Control Scale, Pilowsky’s Depression Questionnaire, Taylor Manifest Anxiety Inventory. The treatment outcome measured was based on weekly measurements of a six-point categorical scale (no pain at all - so intense that it is incapacitating). Regarding short-term outcome, 88% had pain reduction of approximately 50%, Blanchard procedure. After two years, the post-treatment evaluation (long-term success) showed that 42% of the patients were symptom free, 26% were significantly better, 13% were slightly better, and 18% had no change. The best predictors (discriminant functional analysis) were: a) short term success: pain duration (-0.56), depression (-1.14); long term: depression (-0.61) measured by the Cronbach Alpha. In their prediction model, they were capable of predicting success in 83% of the patients in the short-term and 100% in the long-term.

In a study by Strychalski et al. (1984). A telephone survey to assess treatment outcome success was conducted 2-3 years after treatment of patients suffering from TMJ syndrome. The short-term (immediately post-treatment) improvement rate was 81%. Some 35% had complete pain remission. The long-term (2-3 years after treatment) persistence of improvement rate was 72%. The study included 31 individuals who had been previously treated for TMD. All patients were treated by a single clinician. The ages ranged from 18 to 72 years with an average of 39.3 years. There were eight males and twenty-three females. The treatment outcome measure was the phone interviewer assessment of patient’s response to the following question: ‘Was the problem relieved
during treatment? Yes, Partially or No’. Review of the literature for long-term success rate after 1 to 9 years post-treatment (complete remission, partial remission of symptoms): 60-98%.

In a study by Heloe et al (1980), the data obtained from 113 female patients with MPD who had been previously been examined by a multidisciplinary team. After 1 1/2 year the patients were interviewed concerning possible improvement (totally improved, partially improved, not changed, or felt worse), present symptoms and perceived gain from the treatment. Forty-two percent of all patients reported total improvement, 39% partial improvement, 14% no improvement, and 5% got worse.

In a study by Millstein-Prentky & Olson (1979), responses to the Minnesota Multiphasic Personality Inventory (MMPI) were used to develop a 29-item scale designed to predict the treatment outcome of 135 patients with Myofascial Pain-Dysfunction Syndrome (MPD). The population was comprised by 135 patients (20 female and 10 male), average age 34.53 years (range 13-71) with no gender differences in mean age. The treatment outcome measure was evaluated by patient self-report of no symptoms at the end and no more then one pain episode during treatment (62.9% successful). The results show decline in the discriminating ability of the MMPI. This was due to absence of consistent differences in personality between successful and unsuccessful MPD patients. Both groups showed similar profile configurations (psychosomatic-V); however, the unsuccessful group had higher profile elevations.

The composite MMPI profiles of 42 successfully and 42 unsuccessfully treated female patients with MPD syndrome were compared statistically by Schwartz et al.
The profiles did not differ in configuration, but that of the unsuccessful patients was significantly higher than that of the successful group, indicating a greater degree of emotional distress. The pattern of deviation from normal for both groups was diagnostic of a psychophysiological disorder marked by depression and somatization. The aim of this study was to compare retrospectively the MMPI scores of successfully treated versus unsuccessfully treated MPD syndrome patients in order to discover any distinguishing personality characteristics. The outcome measure was a clinician's assessment of absence or significant reduction of all symptoms of MPD after one to three months. The treatments provided were medications, bite-plates, exercises, physical therapy, and placebo. The non-responding patients had an average age of 36.1 years (range 13-59), while the responding was 34.0 years (range 16-69), all women. The results showed that patients differed at the 1% level of significance on the Depression and Psychopathic Deviate scales, suggesting that the major departure from their successful counterpart was in their degree of depression, their agitation or anger. They also differed (p<0.05) in hypochondriasis, hysteria, psychasthenia, and schizophrenia. The predominant finding of this study is that although non-responding patients with MPD syndrome are similar to the responders in terms of their constellation of personality traits, they show a greater overall degree of emotional distress. Moreover, both groups showed certain deviations from normal, suggesting that MPD syndrome patients are generally inclined to hysterical character marked by depression and somatization. The results are similar to patients with low-back pain.
In a study by Small & Hill (1974), fifty patients (26 females, age range 14-54; 24 males, age range 19 to 47) with temporomandibular joint pain-dysfunction syndrome were interviewed and given the Minnesota Multiphasic Personality Inventory (MMPI) and Cornell Medical Index (CMI). According to these instruments, ten patients were classified as normal and 40 patients were classified as abnormal. All patients received the same conservative therapy. Excellent results were obtained in all the abnormal patients but there was a lack of response in seven of the then normal patients. These results indicate that specific personality types could be correlated with TMD treatment outcome.

iii) Neuropsychological variables as predictors

With respect to treatment outcome and psychological status, a retrospective study of post-traumatic TMD (pTMD) patients (Romanelli, Mock & Tenenbaum, 1992) suggested that patients with post-traumatic TMD responded poorly to treatment in comparison to those with TMDs not related to trauma. One of the notable features of the pTMD patients was the high prevalence of symptoms suggestive of affective disorder. It was postulated that the co-existence of affective disorder might be responsible for the non-responding nature of the pTMD patients as compared to their non-traumatic counterparts. Further, these patients also exhibited characteristics which were very similar to those observed in patients with post-concussion syndrome (Stuss et al., 1985). In this regard, patients with post-concussion syndrome have been demonstrated to have a high degree of neuropsychological deficits as determined on the basis of a variety of
neuropsychological tests as well as various degrees of affective disorders. As far as TMD is concerned, just one study has evaluated neuropsychological deficits in patients with idiopathic TMD and post-traumatic TMD (Goldberg et al, 1996). The authors found out that both groups differed in a number of different neuropsychological and clinical features: simple and complex reaction time tests, the Brown-Peterson Consonant Trigram Test, and the California Verbal Learning Test, and reaction to muscle palpation. These similarities between post-traumatic TMD and patients with post-concussion syndrome further characterize and elucidate the possibility that pTMD patients might have neuropsychological deficits (Goldberg et al, 1996) in addition to or as part of the apparent affective component. However, there also appeared to be neuropsychological deficits (in comparison to external normal values) in the non-traumatic group as well although not to as a great a degree as the pTMD patients. Inasmuch as the pTMD patients appear in general to be more refractory to treatment than non-traumatic TMD patients, it was further hypothesized that neuropsychological deficits in addition to other aspects of affective disorder, might also be associated with and perhaps even predictive of poor treatment outcome, not only in pTMD but perhaps in some forms of non-traumatic TMD as well. However, the sample size was small and the study was not designed for answering this question.
iv) Summary

After many years of studying the psychological traits and states of patients and their relationship to the etiology of TMD, it can be concluded that there is not enough evidence to confirm the etiology of TMD by assessing past or current psychological factors for an individual patient. There was no agreement in studies evaluating if there were personality differences between TMD responding and non-responding patients (Schwartz et al., 1979; Millstein-Prentky & Olson, 1979). However, the evidence is strong enough to affirm that psychological factors, particularly depression and somatization, do have an important role in the perpetuation and treatment of MPD syndrome (Gale & Funch, 1984; Gerschman et al., 1987; Fricton & Olsen, 1996). These findings have implications for therapists who wish to understand the disorder and to manage patients appropriately. Clinicians should recognize the importance of nonspecific factors such as placebo effects, doctor-patient interactions, and spontaneous recoveries in the treatment response. This awareness will enhance their effectiveness as therapists and will help them to avoid using excessive or radical treatment methods (Rugh & Solberg, 1976; Greene & Laskin, 1983).

Many studies have established that a substantial proportion of patients with TMD demonstrated psychiatric symptoms. While there is controversy about the causal relationship of pain and psychiatric symptoms (Schwartz et al., 1979; Millstein-Prentky & Olson, 1979), there is the need of studies analyzing the role of psychosocial variables with the perpetuation of chronic pain. In other words, there is still the need of well
designed longitudinal studies with accurate definition of the population and methods employed using reliable and valid instruments in order to assess the role of psychosocial variables in the perpetuation of TMD in a reproducible manner.

In addition, treatment of facial pain would be more effective if response to specific therapy could be predicted. Identifying potential responding and non-responding patients to specific therapy before initiation of care would enable the dentist to develop a treatment plan that would be most helpful and appropriate for a specific patient. If the variety of treatment options for patients with TMD could be narrowed, much unnecessary treatment might be avoided.

As indicated above, some studies have been carried out using psychosocial variables as predictors of treatment outcome for TMD, but further research is still needed to determine whether non-specific treatment factors such as cyclic changes in physical signs or psychosocial status account for the failure to find differential treatment outcomes for specific treatments and providers.

In addition, the available literature is difficult to compare due to different case definitions, duration of the studies, prospective versus retrospective design, treatment outcome measures, treatment regimens and duration as well as different statistical techniques employed. However, the majority of the available literature seems to suggest that psychosocial variables do play a role in predicting the treatment outcome of TMD patients and may be, at least, a perpetuating factor. On the other hand, available literature is limited to a single study in regard to neuropsychological tests, i.e. which measures task
performance in the context of behavior, (Goldberg et al., 1996). Therefore, there is the urgent need for increasing knowledge in this area too.
g) Rationale for pain assessment as the main outcome measure

i) Definition and importance

Pain is an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage. Pain is always subjective. Each individual learns the meaning of pain through experiences related to injury in early life (Magnusson et al., 1995). It is well documented that the prevalence of signs and symptoms of TMD have been as high as 88% in the general population, but only about 5% to 26% in the population samples have been considered to be in need of treatment due to pain. In addition, pain has been shown to be a major reason for patients to seek treatment for any condition (Suvinen & Reade, 1995).

ii) Pain assessment

The above definition implies that, to measure pain, several aspects have to be considered. Several verbal and non-verbal scales have been developed and extensively discussed. Even though many scales have been found to be reliable and sensitive, it is important to emphasize that there is no ‘gold standard’ in quantifying pain (Ekblom & Hansson, 1988; Duncan, Bushnell & Lavigne, 1989; Price et al., 1994). There is however general agreement that it is important that the pain measuring scale should place a
minimal burden on the patient, should be well understood by patients, should provide a wide range of scores, should show sensitivity to analgesic intervention, and should demonstrate the appropriate reliability and validity. The scale should also be appropriate for the patient group being investigated (Kopp, 1977; Harms-Ringdahl et al., 1986).

In laboratory studies using human subjects, the quantification of pain has been by psychophysical methods (thresholds), rating scale methods (visual analogue scales), magnitude estimation procedures (number assignment or hand grip force), and by performance behavior of laboratory tasks (Chapman et al., 1985). Pain experience has been measured by physiology correlates, such as electromyographic measures, autonomic studies, evoked potentials, and electroencephalographic methods. In clinical pain assessment, the most common methods employed have been behavioral measurements, subjective pain reports (such as visual analogue scales of pain) and word descriptors (such as the McGill Pain Questionnaires), as well as pain inventories. In the literature, opinions vary between researchers as to which scales are most sensitive. Some advocate the non-verbal scales, such as the VAS (Price et al., 1986); while others prefer the verbal scales, such as categorical scales for pain improvement (Duncan, Bushnell & Lavigne, 1989). In relation to the above, I would like to make a brief review of the advantages and disadvantages of the most widely used instruments of clinical pain assessment.
iii) Subjective pain reports and word descriptors

Subjective pain reports are the most commonly employed procedures in clinical pain research, and are used to scale both pain itself and pain relief following treatment. In the first case, the patient is asked to report the intensity (or some other feature) of the pain by using a standardized judgment procedure. In the second case, he is told to report the amount of relief afforded by an analgesic treatment. The simplest report instruments are pain category ratings and VAS judgments. Category scales are often problematic. When categories of pain description are employed, it is difficult to specify the size of each category and whether the categories are of equal spacing. VAS methods have achieved a greater popularity than category scales in clinical work just as they have in the laboratory. The efficiency and simplicity of VAS scales are important in clinical research. VAS places minimal demand on sick patients while poorly educated patients can usually grasp the nature of the scale with little difficulty (Chapman et al., 1985).

The visual analogue scale (VAS) is a simple and frequently used method for the assessment of variations in intensity of pain. In clinical practice the percentage of pain relief assessed by VAS, is often considered as a measure of the efficacy of treatment (Carlsson, 1983). The Visual Analogue Scale (VAS) consists of a 100 mm line with the ends defined as 'no pain' and 'worst pain imaginable' (Magnusson et al., 1995). The patient registers the degree of pain with a pen (X mark) on the chosen part of the line. The pain is expressed as the distance in milimiters from the end point to the pencil mark. The sensitivity and precision of VAS has been extensively studied. Sensitivity of the scale is
defined as the capacity of a scale to demonstrate a change in the pain intensity of the degree of discomfort. A scale with a great change between recordings is considered to be more sensitive than a scale exhibiting a small change. Precision is defined as the absence of haphazard errors. A scale is judged to have good precision if the same value obtained at repeated recordings. In most analysis of the VAS, the actual figure +5 mm was used, namely, a difference +5 mm between two recordings was ignored.

However, such scales are sometimes problematic. For example, Carlsson (1983) provided evidence that the reliability of VAS is weak. In addition to response bias, the major problem of VAS methods is poor sensitivity to treatment effects. In some cases, affective states may confound assessment with these and related techniques by inducing a response set. A more basic problem is that unidimensional scales oversimplify the complex human experience of pain. One alternative is scaling based on word descriptors, or verbal scales, such as the McGill Pain Questionnaire.

The McGill Pain Questionnaire has been used widely and provides a powerful way of obtaining data on both the quantitative and qualitative aspects of pain. The fact that it scales pain multidimensionally is a major advantage. One limitation of the MPQ is that patients sometimes have difficulty with the complexity of the vocabulary it uses. If the words presented are too demanding for the patient, his or her compliance with test administrator’s instructions cannot be assured. Even if the patient is well educated, the test may reflect in part the vocabulary limitations of the patient as well as the nature of the pain. Because of these problems, the MPQ may prove problematic when comparisons are made across cultural or subcultural groups. In such studies, cultural differences in
language habits may be confounded with differences in pain expression. In addition, it is
time consuming to administer than VAS instruments and category scales (Gagliese &
Melzack, 1997). Finally, the MPQ weights sensory aspects of pain more heavily than
affective and evaluvative. Because of that, patients give more consideration to the sensory
aspect of pain than to the affective or evaluative aspects in testing process, and this biases
the results (Chapman et al., 1985). However, Gagliese & Melzack (1997) have shown
that both affective and sensory aspects of the MPQ correlate very well with each other not
only in young adult but also in the elderly population.

Magnusson et al. (1995) tested five different scales of self-assessment of pain in
patients with temporomandibular disorders. The precision and sensitivity and the capacity
to register memory of pain and discomfort were compared for each of the five scales.
Two non-verbal scales (the visual analogue scale and the numerical scale), two verbal
scales (Behaviour Rating Scale and Verbal Scale) as well as a combined one (Combined
Scale) were tested. The ability to recollect the degree of pain (pain memory) and
discomfort with the different scales was tested by comparing the first and fourth
recordings. The first recording was at the first visit, the second at 2-3 weeks after the first
visit, the third recording was at 8-10 weeks, and the fourth recording was one year after
the start of treatment. A high correlation between these recordings would mean that the
scale is suitable for use when measuring memory of pain and discomfort. The authors
used the Behaviour Rating Scale (BRS) which is a 6-point scale where the steps are
expressed in sentences such as 'Minor pain/discomfort, only noticed when I think of it'
and 'Very strong pain/discomfort, totally handicapping, cannot do anything because of
the pain/discomfort'. Magnusson and colleagues found that more than one-third of the patients (35.7%) had their symptoms for more than ten years when they came to the department, and only one had a history of symptoms for less than six months. Eighty nine percent of the patients experienced their symptoms daily, and fifty one percent had previous treatments with no or insignificant treatment success. The mean values and range for the whole group of patients according to the five scales decreased between recordings. When comparing the differences between the first and fourth recordings (initial visit and one year after), the BRS (Behaviour Rating Scale) had a higher precision (Kappa 0.68, p<0.0001) when measuring pain than the VAS (Kappa 0.38, p<0.003). Regarding sensitivity, both VAS and BRS showed that the discomfort caused by the subjective symptoms had decreased after 8-10 weeks of active treatment (p<0.002 and p<0.001, Wilcoxon, respectively); however, only the BRS showed statistical significance (Wilcoxon, p<0.002) when measuring changes in pain levels pre- and post-treatment. The BRS was superior to other scales when measuring changes in both pain and discomfort. When recording the memory of pain and discomfort one year after start of treatment, the BRS exhibited the best result (Kappa 0.63, p<0.0001 and 0.39, P<0.01, respectively).

Regarding the precision of the scales, all showed decrease in pain and discomfort from the first to the third recording, even in 21 patients who received no active treatment. It is well known that symptoms of TMD fluctuate spontaneously with time (Kopp, 1977), but a more likely explanation for the decrease is that all patients were given counselling regarding TMD in their first visit. Both BRS and VAS were the only scales that showed acceptable kappa-values for pain and discomfort. Magnusson's results contradict previous
investigations where high precision has been found for the VAS (Ekblom & Hansson, 1988; Price et al., 1994). However, most of these studies mentioned dealt with either experimental or acute pain while the present patients had chronic pain. Magnusson et al., 1995, also found that when the precision of a scale is being estimated, a scale with few steps is more favourable than a scale with more steps since the possibility of agreement by chance is higher. A negative linear correlation was found between the degree of correlation and the number of steps in these scales.

A low precision for the VAS and reports of difficulties in understanding and using this scale have been presented earlier (Carlsson, 1983). The author found out that the validity of VAS estimates performed by patients with chronic pain may be unsatisfactory. Two types of VAS, and an absolute (pain intensity) and a comparative scale (pain relief), were compared with respect to factors influencing the reliability and validity of pain estimates. A pain relief scale is a ‘comparative’ visual analogue scale, the ends of which are defined in terms of degree of pain relief. It may have several advantages as compared to an ‘absolute’ analogue scale as all patients start at the same baseline; a comparative scale may also be easier to use. As shown in this study, the absolute type of VAS seems to be less sensitive to bias than the comparative one and is therefore preferable for general clinical use. Repetitive use of the scales may influence the correlation between the two types of estimates of pain intensity. Patients may learn the use of the scales, which should result in a higher correlation. Subjective evaluation of change in pain necessarily involves the recollection of the level of intensity on a previous occasion. Hunter et al. (1979) have found that the memory of pain was unexpectedly accurate when verbal pain descriptors
were used but it is not known whether this also applies to other methods of assessment. Moreover, the patients appear to differ considerably in their ability to use the VAS reliably. When assessing efficacy of treatment, attention should therefore be paid to several complementary indices of pain relief as well as to the individual's tendency to bias his estimates.

The mean correlation between the two types of analogue scales was significant, but the correlation between the two scales was low when pain is indicated as decreasing, as compared to unchanged or increasing pain. The magnitude of the mean correlation between the two forms of VAS is comparable to that reported by Elton et al., 1979, and Scott and Huskisson, 1976, in which studies VAS and other scales were compared for the evaluation of treatment of pain. However, in a number of other studies higher correlations were found between VAS and numerical and verbal rating scales (Downie et al., 1978; Ohnhaus & Adles, 1975; Woodforde and Merskey, 1972). These studies used large groups of patients who estimated pain intensity using several scales, all administered in one occasion. Such a design is usually chosen in order to reduce the systematic bias due to interindividual variability. Magnusson, List & Helkimo (1995), measuring pain before and after TMD treatment, found high correlations between two of the five different self-assessment scales, the VAS and the Behaviour Rating Scale regarding precision and sensitivity.

Price et al., 1994, compared numerical rating scales and mechanical visual analogue scales (M-VAS) for their capacity to provide ratio scale measures of experimental pain. Separate estimates of experimental pain sensation intensity and pain
unpleasantness were obtained by each method, as were estimates of clinical pain. Orofacial pain patients (n = 23) diagnosed with myofascial pain dysfunction (21 women, 2 men, mean age 38.7 years ± S.D. = 19.5) made numerical scale and VAS ratings in response to noxious thermal stimuli (45-61°C) applied for five seconds to the forearm by a contact thermode. The derived stimulus response function was well fit as a power function only in the case of sensory M-VAS. The power function derived from sensory M-VAS ratings predicted temperatures chosen as twice as intense as standard temperatures of 47°C and 48°C, thereby providing evidence for ratio scale characteristics of M-VAS. The stimulus-response function derived from sensory numerical ratings differed from that obtained with M-VAS and did not provide accurate predictions of temperatures perceived as twice intense as 47°C or 48°C. Both M-VAS and numerical rating scales produced reliably different stimulus response functions for pain sensation intensity as compared to pain unpleasantness and both provided consistent measures of experimental and clinical pain intensity. Finally, both mechanical and pencil-and-paper VAS produced very similar stimulus-response functions. The author concluded that both VAS and M-VAS were adequate for use in both clinical and basic research.

Ekblom & Hansson, 1988, compared six different pain rating scales, including a "pain relief scale", in 80 randomly assigned orofacial pain patients (36 males and 44 females, aged 18-68 years) from an emergency clinic for dental and oral surgery. The patients had suffered from pain for one to four days. The causes for their pain were either pulpal inflammation, apical periodontitis, pericoronitis or post-operative pain following removal of a tooth. Pain intensity measurements were made before and after a
thirty minute period of afferent stimulation (TENS/vibration and placebo). A good correlation was found between pain scores derived from pain relief scale, visual analogue-, numerical- and graphic-rating scales (Pearson product-moment correlation coefficient, range 0.89-0.98, p<0.001) for both measurement in mm or change in pain intensity expressed as percentage. The verbal rating scale did not perform well. Using linear regression, the slope of the different lines varied between 0.91 and 1.00. According to the authors, the pain relief scale and the numerical rating scale are interesting alternatives to the established visual analogue scale.

Duncan, Bushnell & Lavigne, 1989, compare verbal descriptor scales (VDS) and visual analogue scale (VAS) for pain assessment. Eight subjects (8 healthy drug-free pain volunteers, 4 males, 4 females, age range from 21 to 26, French speaking) rated painful and near-painful heat stimuli (42° to 51° C, 5 sec pulses to 6 spots of skin, 3 on either forearm) by using visual analogue scales for intensity or unpleasantness and by choosing the most appropriate phrases from lists of intensity or unpleasantness descriptors in 4 sessions. In the intensity dimension, the relationship between perception and stimulus temperature was essentially identical whether calculated from the visual descriptor scales (MANOVA, p > 0.93). In the unpleasantness dimension; however, the curves derived from the VAS and VDS methods diverge, but this difference was non-significant (p = 0.17). Alternatively, data derived from the verbal descriptor scales revealed that subjects rated the painful temperatures as relatively more intense than unpleasant (p = 0.006); this difference could not be detected using the visual analogue scales (p = 0.08). In addition, differences between end-phrases ‘extremely severe’ and ‘worst pain imaginable’ were
very small when compared to the mean rating of the VAS (95.8% versus 100%, respectively) for pain intensity and (86.5% versus 100%, respectively) for pain unpleasantness. These results confirm that both visual analogue and verbal descriptor techniques successfully quantify sensory intensity and affective aspects of pain, but that verbal descriptors may provide the more sensitive tool for separating intensity and unpleasantness.

Harms-Ringdahl and coworkers (1986) compares intensity levels assessed on Borg’s Category Scale for Ratings of Perceived Pain - BRPP (a verbal scale using adjectives and adverbs combined with the number 0-10) with assessments on the Visual Analogue Scale - VAS (a 10cm horizontal line). Eight healthy subjects (males, mean age 27 years, range 24 to 31) volunteered in an experimental study, where pain was provoked by load on passive soft tissue elbow joint structures. Each subject participated 4 times on different occasions in the same experimental set-up, which was divided into six 2 min periods; 3 periods with load induced by applied external weights causing load moments of 3.4 Nm, 4.5 Nm and 6.8 Nm plus that induced by the weight of the lower arm and hand (average 2.9 Nm), followed by 3 periods without external weights. Each series consisted of 12 assessments given during the last 10 sec period of each minute on either the BRPP or the VAS. No significant difference was found between the first and second time a scale was used by the same subject, and none between the assessments on the VAS and the BRPP (correlation coefficients between 2 scales ranged from 0.78 to 0.99, with a mean of 0.90). Intensity levels of pain increased with load and time and decreased after reduction of the load moments. It is concluded that both scales can be used to reliably assess
intensity levels of perceived pain elicited by loading joint structures. Intensity levels, as assessed on both scales, are associated with applied external load and time for exposure.

One study analysed the different types of VAS regarding precision. Seymour et al. (1985) compared the accuracy of visual analogue scales (VAS) of different lengths (5, 10, 15 and 20 cm) and same boundaries (no pain and worst pain imaginable) and with different end-phrases (troublesome, miserable, intense, unbearable and worst pain imaginable) were used to record pain in 50 male and 50 female patients (18 years or older) with pulpitis or pericoronitis (acute pain). All 100 patients successfully completed the self-administered questionnaires. High correlation was found between the scores on all the scales (0.56 to 0.91). Scales of length 10 or 15 cm had the smallest measurement error. The scale with end-phrase ‘worst pain imaginable’ was found to be the best choice for comparing present pain or worst pain between different groups. Using this scale, no significant difference was found between the scores of males and females or between those of patients with pulpitis and pericoronitis. This study suggests the use of 100 mm VAS with the end-phrase ‘worst pain imaginable’ as being the most suitable for measuring dental pain.

Most studies regarding pain assessment were done in young adults, and very few in the elderly population. Higher frequency of incorrect responses to the VAS has been associated with the patient population; however, in a study by Gagliese & Melzack (1997), the VAS and the Verbal Descriptor Scale (VDS) were significantly correlated with each other. In addition, 40% of the patients felt that the VDS was the easiest to use
and described their pain best. These data suggest that both VAS and VDS may be used effectively in the assessment of chronic pain in the elderly.

iv) Memory for pain

Assessment of pain and discomfort based on memory plays an important part in the evaluation of treatment results in clinical work. Examiners often ask patients to estimate their pain and discomfort compared with pre-treatment levels. According to several previous studies, the memory of pain is modest (Erskine, Morley & Pearce, 1990), and it has also been found that patients tend to overestimate their pre-treatment pain levels (Linton & Melin, 1982; Linton & Gotestam, 1983). This is in line with the findings of a study (Magnusson et al., 1995) where the average pre-treatment pain was higher for the five scales one year after treatment. Of the five scales, only the BRS and the VS reached statistically significant levels for memory of both pain and discomfort; they were also the only scales that reached acceptable, or close to acceptable, Kappa values. Of these two scales, the BRS was superior to the VS. The superiority of the BRS when recording memory of pain has also been put forward by others (Linton & Gotestam, 1983). The validity of the pain scales were not determined due to a lack of a reliable gold standard. Some authors have have assessed a scale validity by using the VAS as the gold standard (Linton & Gotestam, 1983), or by comparing the scale to a physically measurable stimuli (Price et al., 1986, Price et al., 1994). When working with pain patients, one must,
however, always bear in mind that 'the central point is that the patient reporting pain must be evaluated, not just the pain itself (Turk, 1993).

Linton and Gotestam (1983) studied fifteen back or joint chronic pain patients (7 males, 8 females, mean age = 43 years, pain duration for 1 to 26 years) rated their pain intensity of both a visual analogue scale (100 mm) and verbal scales (0-5 point Verbal Pain Scale) so that comparisons between the scales could be made for each subject. Subjects made a pre-baseline estimate of their pain and then they rated their pain throughout a baseline and treatment period averaging 5 weeks. Four to 9 weeks after baseline, subjects were asked to remember how much pain they had at baseline. Discrepancies were noted on both scales; two-thirds of the subjects had discrepancies which were larger than 10% of their baseline rating on the VPS scales, while 87% had such discrepancies on the VAS. The mean discrepancy, regardless of direction was 15% for the VPS and 23% for the VAS. Thus, the VAS produced significantly greater discrepancies than the verbal scale; however, discrepancies in the verbal scale were in both directions (overestimation, underestimation) while in the VAS were only unidirectional (overestimation). Results indicated that two-thirds of the individual subjects had significant correlations between the scales with a mean of 0.68 (range 0.29 to 0.96). The one-third of subjects who did not have significant correlations also had significantly less variability in their daily pain ratings (pre-baseline estimation) than did subjects with significant correlations; demonstrating less compliance. This may help to explain the lower correlation found in this particular group.
Twelve chronic pain patients (6 males, 6 females, mean age = 48) were employed in an investigation by Linton & Melin (1982) of the accuracy of memory for chronic pain. All subjects had pain for more than 6 months duration (mean 2 years). The primary site of the pain was in the back or joints, and pain was measured using 100 mm VAS. One is generally skeptical of the accuracy of post-hoc ratings, since they may be influenced by what happens during the interim between experience and recall. Longer recall periods are more relevant; however, in the chronic pain situation because pain or its treatment ordinarily continues for at least several weeks. Furthermore, in the evaluation of a treatment program, the treatment experienced may systematically influence the remembering of pre-treatment pain levels. Subjects first made pain ratings before entering a treatment program. At dismissal, 3-11 weeks later (mean = 51 days, range = 24 - 75 days), they were asked to remember how much pain they had had at baseline. Results show that patients remembered having significant more pain than they actually rated during the baseline period. Caution is therefore warranted when using post-hoc pain measures with chronic pain patients. Of the 12 patients, 11 (92%) remembered the pain as being more severe than their actual baseline rating. The mean actual baseline rating was 56, while the recalled rating was 69. This 19% difference was statistically significant (Wilcoxon test, p < 0.01)(Carlsson, 1983).

One of the explanations for the discrepancy between chronic and acute pain patients is the fact that memory for pain is accurate if it is related to a specific event (e.g. acute) and if the recall period is relatively short; while memory for pain is less accurate if the pain is long standing and the recall period is long (e.g. chronic). Regarding this issue,
the author's results also seem contradictory. A high Spearman's rank correlation (0.53, p < 0.05, one-tail) between actual baseline ratings and remembered baseline rating was found; which contrasts with the differences reported earlier. The author tried to justify this contradiction stating that "the degree of improvement systematically affected recall ratings"; however, the correlation between error size and amount of improvement (actual baseline pain minus dismissal pain) was not significant (r = 0.23, N.S.). One possible explanation given was that "remembering was related to the recall interval"; nevertheless, the correlation between error size and interval length was not significant (r = 0.06, N.S.).

Finally, the sample size was small (n = 12) which compromises the external validity of the study; the population had back and joint pain, which is more debilitating than TMD, and may have affected the results (Von Korff et al., 1988); and the average age was higher than reported for the TMD population (48 versus 30, respectively) which may also have affected the results.

In a review paper, Erskine et al. (1990), reported a qualitative review of the literature on memory for pain. Most research has focused on the accuracy of memory for pain intensity. There is some evidence that recall is moderately accurate but this conclusion is tentative because of significant methodological problems. There is also some evidence that recall of acute pain is more accurate than recall of chronic pain. Patients with chronic continuous pain had a discrepancy varying from to 2 to 23% between baseline measures and recall, which a tendency to overestimate baseline pain intensity on the VAS. Most studies were done in patients with heterogeneous populations
(joint, low back, headache and cervical back chronic pain) and variable sample sizes (10 to 93 patients) which may have affected the results.

Another interesting finding is that current pain state influences estimates of past pain in a systematic manner. This memory bias effect is independent of changes due to treatment outcome or learning occurring during therapy. Therefore, the outcome seems to influence pain intensity assessment and not the opposite. It should be noted that this finding pertains only to memory for pain intensity. Memory for the affective, sensory, behavioural, situational and other contextual aspects of pain is unexplored (Eich et al., 1985). One study tried to assess memory for pain quality by comparing the sensory and reactive scale of the McGill Pain Questionnaire (MPQ): total MPQ score and the Number of Words Chosen (NWC) measure from the MPQ, respectively. Fourteen patients with chronic rheumatoid arthritis, who were tested for ratings pre- and one week after surgery, were compared to 23 healthy individuals. The different sources of pain, natural versus induced, make the interpretations difficult because the source of the pain may have influenced the results (e.g. natural pain may be better recalled than experimental pain, or vice-versa), but the correlations for the two subscales were non-significant, suggesting poor recall of the quality of pain.

Higher frequency of incorrect responses to the VAS has been associated with the elderly population; however, in a study by Gagliese & Melzack (1997), the VAS and the Verbal Descriptor Scale (VDS) were significantly correlated with each other. In addition, 40% of the patients felt that the VDS was the easiest to use and described their pain best.
These data suggest that both VAS and VDS may be used effectively in the assessment of chronic pain in the elderly.

v) Pain Assessment in TMD

The subjective and private nature of pain experience means that it can be measured only indirectly, i.e., by how it is described by the patient or by observing the patient's behavior. Patients with pain generally form a heterogeneous population. Most present with differing disease status; e.g., the intensity of pain may vary, not only between, but also within individuals at different times (Raphael & Marbach, 1992). It is now recognized, however, that pain, including pain in TMD, is a complex, multifactorial experience including not only sensory-discriminative dimensions, but also motivational, affective, and cognitive factors that all interrelate and affect the pain response and expression by the patients. There has been few studies that have used validated instruments or psychometrically tested the validity and reliability of the instruments used in patients with TMD (Fricton & Olsen 1996; Gerschman et al, 1987; Gale & Funch, 1984). Many have lacked longitudinal designs; e.g., it is not known how different psychological dimensions change with time. There has been a lack of evidence for specific psychologic profiles for patients with TMD. However, the studies have indicated a multifactorial basis for TMD, and thus a need for the assessment of patients with TMD to be from a multifactorial perspective. Few studies exist where multiple factors have been studied systematically. Multivariate statistical principles have been applied only
recently to analyze the contributions of multiple variables in factors associated with patients suffering from TMD. The selection, diagnosis, and exclusion criteria, as well as the drop-out rate or intend-to-treat groups should be well documented.

The measurement of pain in patients with TMD has been generally based on an objective assessment of pain elicited on palpation of the muscles of mastication or the TMJs. Many indices have been constructed to ascertain a severity value for the responses obtained. Of these, the Research Diagnostic Criteria for Temporomandibular Disorders provides a comprehensive assessment of TMD signs and symptoms while at the same time emphasizing the potential role of psychologic variables in mediating the pain response (Dworkin & LeResche, 1992). The TMJ scale, which is a self-report symptom inventory, tests for the clinical significance of pain report and joint dysfunction and has been found to have predictive value in detecting psychologic problems in TMD patients (Levitt, 1990). The subjective pain experience has been measured by anamnestic (history) indexes, symptom severity indexes, or by visual analogue scales.

vi) Other treatment outcome measures used in patients with chronic pain

There has been a diversity of outcome criteria used to establish treatment success in chronic pain treatment. Physical functional assessment, global ratings by health care providers, use of health care system, return to work rates, reduction in pain behaviors, self-reports of activity, self-reports of medication use, self-reports of pain intensity, and
psychological measures of distress or psychopathology. There is no correct efficacy measure but rather each must be looked at within its appropriate context. For example, third-party payers may focus on whether the patient has returned to gainful employment. For the clinician and the patient, pain reduction and improved quality of life may be the favored outcome. For health care providers, reduction in the use of the health care system may be the favored outcome. Different populations studied and different treatment outcome criteria utilized make the results impossible to be reproduced and the studies impossible to be compared, even with meta-analysis techniques (Turk et al., 1993a). One of the possible solutions is to use multiple outcome criteria which are correlated (Minnesota Multiphasic Personality Inventory, McGill Pain Questionnaire, and Beck Depression Inventory) or to compute component scores among statistically related measures derived from principal component analyses and determine the treatment efficacy based on patient changes on component scores. The advantage is that multiple component scores are more reliable than individual scales; however, the sample size required to perform such analyses are larger than usually used in pain outcome studies. In addition, the use of aggregate scores creating several categories of outcome measures still leaves open the possibility of differential outcomes on the various components.

In order to improve the reproducibility of the studies, the following aspects must be taken into account when establishing a criteria of success of treatment outcome: a) precise definition of populations studied (inclusion/exclusion), b) precise description of treatment provided; c) precise definition of outcome, and d) precise definition of successful treatment outcome; e) use of reliable outcome measures; f) the effects that
different treatments may exert on different patients; g) treatment costs (financial, emotional) and efficiency; and h) percentage of patients available at follow-up; i) durability of treatment success; j) acceptability of the treatment by the patient; h) significance of change for the patient, professional, society and third-party payers; and k) clinical versus statistical significance of chances following treatment.

Colvin et al., 1980, noted that the majority of patients indicated that they desired total and permanent relief of pain as their primary goal. Specifically, 97% of the patients (n = 287) indicated that they would accept a 50% improvement and 52% would try elsewhere if the treatment did not bring acceptable pain relief. Duckro et al., 1985-86, and Newman et al., 1978, reported significant improvements in such measures as reduced use of medical resources and prescribed analgesics, improvement in measures of physical function, and increased ability to cope with pain. Wang et al., 1980, conducted a 1 to 3-year follow-up study of patients treated in an outpatient pain clinic. Of an initial sample of 725 patients treated, 361 (50% response rate) responded to a questionnaire that asked them to indicate their belief as to whether the treatment had been beneficial to them, and 52% still complained about the problem. Chaplin, 1991, reviewed a number of treatment outcome studies and concluded that although these studies show a statistically significant decrease in self-rated pain intensity, the magnitude of reduction tended to be small, ranging from 10 to 20%. Carron et al., 1985, reviewed studies of patients treated at pain rehabilitation programs in both New Zealand and the United States and noted that minimal reduction in pain was reported at the one-year follow-up. In both countries, 39 to 45% of the patients indicated that they considered themselves at least 'somewhat'
improved. Cassisi et al. (1989) reported 61% of patients who had completed the treatment program reported a statistically significant (27%) decrease in pain. They noted that despite the relatively modest reduction in pain, 79% of the group reported that they were satisfied with the way they were treated in the program. In addition, 74% said that they would recommend the program to a friend who had a similar condition, suggesting good correlation between patient’s self-assessment and visual analogue scales. The author also found out that patients who completed treatment reported significantly fewer physician visits than those who were not approved for treatment by insurance carriers, and that treated patients had significantly fewer surgeries following treatment than patients who declined to participate in treatment. They also reported that 69% of program completers had returned to work by follow-up as compared to 41% and 39% return to work for those patients who declined treatment by insurance carriers and patients who declined treatment on their own, respectively. The percentages can readily be converted to dollars saved. These data suggest that patients may use criteria other than pain reduction alone to evaluate the treatments received; however, pain reduction correlates very well with other measures of improvement.

Related to the importance of change is the consideration of clinical versus the statistical significance of treatment outcome. When considering treatment efficacy, investigators might wish to view the magnitude of improvement not only compared to pre-treatment levels and other chronic patients receiving the same or alternative treatments, but also to non-pain individuals (Jacobson et al., 1984). How similar are patients’ function levels following treatment to ‘normal’ levels of function in non-
patient? Rudy et al., 1990, demonstrated that at discharge from a pain rehabilitation program, patients approached comparable levels of functional capacity as non-pain subjects. This approach to evaluating treatment success does not categorize patients as 'much improved', 'very much improved', and so on, which commonly has been done in the literature. Rather, the extent to which treatment restores adequate and acceptable levels of functioning can be assessed directly. One strategy that can be adopted in evaluating the importance of treatment is to compute an 'effect size' for each outcome variable. This permits the investigator to determine how much the individual has improved as compared to for example, a control group. A potentially valuable alternative is to compute a reliability of change (RC), (Jacobson et al., 1984; Christensen and Mendoza, 1986):

\[
RC = \frac{x_2 - x_1}{S_{diff}},
\]

where \( x_1 \) is the patient's pretreatment score, \( x_2 \) is the post-treatment score, and \( S_{diff} \) is the standard error of the difference between the two scores, which provides a measure of the spread of the distribution of change in scores that would be expected if no actual change in scores occurs. \( S_{diff} \) can be computed from a measure's standard error of measurement (\( S_E \)) by the formula:

\[
S_{diff} = \sqrt{2(S_E)^2},
\]

with \( S_E \) defined as

\[
S_E = S_t \sqrt{1 - r_{xx}},
\]
where $S_1$ is the standard deviation of a pre-treatment experimental group or a control group, and $r_{xx}$ is the test-retest reliability of the measure being used to determine treatment change. Values of RC that exceed 1.96 are unlikely to occur ($p < 0.05$) unless an actual change in scores occurs between pre-treatment and post-treatment assessment. That is, changes that exceed this magnitude can be considered as reflecting more than the normal measurement fluctuations that occur with repeated testing with a measure that has less than a perfect reliability.

vii) Summary

Pain is the most common reason for patients to seek treatment and must always be assessed, with or without other outcome variables, when evaluating treatment outcome (Suvinen & Reade, 1995). Other outcome variables may also be used depending on the target population to be studied: i) the patient, ii) the clinician, iii) insurance companies, among others (Turk et al., 1993a). However, the pain experience has sensory, affective and cognitive components and it is difficult to be assessed. Many verbal and non-verbal instruments have been developed and are useful in pain assessment. In addition, they usually correlate well with each other (Carlsson, 1983). There is no perfect means of assessment, and all instruments have their advantages and disadvantages (Magnusson et al., 1995). Unidimensional assessment using verbal (such as the Verbal Descriptor Scale - VDS)(Gagliese & Melzack, 1997) and non-verbal scales (such as the Visual Analogue Scale - VAS)(Price et al., 1994) are easy to use but provide limit information concerning
the pain experience. On the other hand, complex instruments like the McGill Pain Questionnaire provide multidimensional information regarding improvement in pain aspects, but are time consuming and require larger sample sizes (Chapman et al., 1985).

Caution must be taken when assessing pain improvement by reduction in baseline assessment using VAS. Pain memory is less precise for chronic pain patients than those with acute or experimental pain. Chronic pain patients usually remember initial pain intensity at follow-up higher than observed at baseline (Carlsson, 1983). However, this per se does not invalidate the VAS as a treatment outcome measure, because the variation has been reported not to exceed 23% improvement (Erskine et al., 1990). Therefore, when assessing improvement, this variation in response must be taken into account besides the variability in the measurement instrument itself. One interesting approach would be to have more than one scale (e.g., one verbal and one non-verbal) in order to have a standard for comparison on how reliable and valid are our pain assessment instruments for that particular population. Another alternative would be to assess chronic pain patients with two different scales at repeated times during the course of treatment assessing their present pain levels. Nevertheless, according to Carlsson (1983), repetitive use of different pain scales may influence the correlation between the two types of estimates of pain intensity. In addition, the patient may learn the use of the scales, which should result in a higher correlation. Therefore, repetitive use of pain scales may bias the results.

As far as TMD treatment outcome is concerned, which usually relies on pain assessment as the main outcome measure, most studies usually report a success rate
which ranges from 60 to 90% regardless of the type of measurement used (Greene & Laskin, 1983). One might wonder that this variation may be due only to methodological differences in pain assessment. However, it is very difficult to attribute this consistent level of improvement only to variation in pain assessment and to the variation in pain memory itself. Therefore, pain assessment in TMD is justifiable and many different valid and reliable approaches have been developed and the choice must be made according to the different objectives and methodologies of each individual study.
II. STATEMENT OF THE PROBLEM

a) Problem 1

Temporomandibular disorders (TMD) are characterized by a variety of symptoms including facial pain which is often exacerbated by jaw movements and particularly by chewing. The most prevalent, painful, orofacial conditions are musculoskeletal in origin and of these, temporomandibular disorders (TMD) are overwhelmingly the most common. Interestingly, it appears that irrespective of the management modality used, 2 to 30% of patients do not improve and may in fact be non-responding to the management approach (Greene & Laskin, 1983).

Despite the high success rate, between 70 to 98% (Greene & Laskin, 1983), one of the major problems in managing temporomandibular disorders (TMD) is the failure of some patients to improve with reversible treatments due to factors ignored or overlooked during the clinical examination. This failure has usually been attributed to the combination of behavioral, psychological and psychosocial factors associated with TMD and their interaction with pathophysiologic factors (Fricton et al., 1988).

Most of the studies on neuropsychological tests have focused on patients with different psychiatric disorders (e.g. schizophrenia) and many other CNS-related conditions (e.g. Alzheimer's disease and closed head injury). As regards to facial pain, only one study has utilized neuropsychological tests in a TMD population. The study by Goldberg et al.
(1996) demonstrated that patients suffering from post-traumatic TMD (pTMD), i.e. patients having TMD symptoms as a result of a Motor Vehicle Accident (MVA), have significantly increased prevalence and magnitude of cognitive and neuropsychological deficits in comparison to patients with non-traumatic or idiopathic TMD (iTMD). This findings are very interesting, because it has also been shown the treatment success rate differs between the two groups (48% for the pTMD vs 80% for the iTMD)(Romanelli, Mock & Tenenbaum, 1992) suggesting that these test could be used as predictors of treatment outcome for the pTMD population. Another interesting finding was that the scores in the pTMD population were comparable to those of patients with closed head injury (Stuss et al., 1985, 1989a, 1989b) at least suggesting that these two populations might have similar etiologies. This finding is very significant, because it provides mechanisms to differentiate among different TMD populations and as an aid for diagnostic classification. Finally, it also helps in the development of appropriate multidisciplinary managements for different TMD groups. Therefore, there is the need to increase the number of studies in dentistry dealing with neuropsychological tests both as disease predictors, and as a tool to improve diagnostic classification and multidisciplinary management.
b) Problem 2

Data have been accumulating that TMD is a chronic pain condition that shares many features with other common chronic pain conditions (e.g. headache and back pain) (Von Korff et al, 1992). Consequently, if neuropsychological tests are shown to be valid as a disease predictors as well as to be comparable between two unrelated chronic pain conditions, they will provide not only valuable clinical information regarding disease classification, but also solid reproducible evidence in favor of the biopsychosocial model.
III. OBJECTIVES

1) The primary objective of this study is to determine the clinical utility of neuropsychological and cognitive tests as predictors of treatment outcome. If this proves to be true, such tests might be used to assess TMD patients prior to management in order to predict those cases in which good outcome might be expected (responding TMD patients, rTMD) versus those in whom a poor treatment outcome will occur (non-responding TMD patients, nrTMD). This probably would save patients and clinicians from the frustration of unsuccessful outcomes and reduce the burden of their unsuccessful managements on the health care system.

2) The secondary objective of this study is to determine if nrTMD patients are similar to another unrelated (not musculoskeletal in origin) functional chronic pain condition, irritable bowel syndrome (IBS). This would provide us with valuable information regarding the most adequate approach to the management of chronic functional conditions such as in nrTMD patients and those with irritable bowel syndrome (IBS). If the two unrelated conditions prove to be similar on the basis of neuropsychological assessment, there is the possibility that they might have similar underlying pathophysiological mechanisms. Consequently, they would be better managed by a multidisciplinary team rather than the traditional search for an unknown underlying etiology, i.e. search for the ideal treatment modality.

3) Our final objective is to assess if the traditional signs and symptoms of TMD can be used as predictors of treatment outcome.
IV. HYPOTHESES

"TMD patients who score higher in the neuropsychological tests will have a negative treatment outcome than TMD patients with normal test values with reversible management strategies for TMD."

If this hypothesis is proved to be true, then a second hypothesis will be tested:

"Patients with irritable bowel syndrome will have similar neuropsychological test scores as patients with nrTMD and that the latter two patient group scores will be different (worse) from patients with rTMD."
V. RESEARCH DESIGN AND METHODS

a) Experimental design

This study was a longitudinal clinical study, six-month follow-up, on the clinical, psychosocial, and neuropsychological characteristics as predictors of TMD treatment outcome. In addition, a comparative analysis was made comparing the neuropsychological features of patients with nrTMD symptoms with those with irritable bowel syndrome (IBS) as well those who responded to reversible management (rTMD).

b) Treatment outcome measures

Baseline measures of pain intensity at rest were made in order to assess treatment effects. Pain intensity at rest is the a major reason for patients to seek treatment and considering that my major goal was to assess pain improvement, this was selected as the main outcome measure. The limitations of this approach is that this is a unidimensional assessment of a multidimensional experience (Suvinen & Reade, 1995). Visual analogue scales were selected due to their efficiency, reliability, validity and simplicity (Chapman
et al, 1985; Price et al., 1986, 1994). The VAS was a 100 mm continuous scale with the extremes anchored in the following statements:

How Severe is your facial/jaw pain at rest within the last month?

| No Pain | Extremely Severe Pain |

A 100 mm scale was chosen, because it has the smallest measurement error when compared to 50, 150 and 200 mm scales; however, it must be pointed out that all scales correlate well (0.56 to 0.91). In addition, the end-phrase 'extremely severe' was found to be one the best choices for comparing present pain and worst pain between different groups, along with the end-phrase 'worst pain imaginable'. Using this scale, no significant difference was found between the scores of males and females or between those with acute dental pain (Seymour et al., 1985). However, there are no studies comparing different types of VAS for chronic pain populations. The question was made asking about the average pain level for the last month, instead of present pain level, in order to prevent daily and monthly variations in pain intensity (Fokard, 1976; Raphael & Marbach, 1992).

In addition, a verbal categorical scale (better, the same, worse) was used for comparison with the VAS:

What do you think about the effect of the treatment in your average pain?

( ) Pain got better   ( ) Pain is still the same   ( ) Pain got worse
The major shortcoming of verbal scales is that because they are categorical, it is difficult to specify the size of each category and whether the categories are of equal spacing. In other words, they are less sensitive to changes in pain intensity than the VAS. This comparison was made in order to assess if there were major discrepancies between verbal and non-verbal scales (Magnusson et al. 1995) in order to increase the validity of the results when comparing them to patient’s self-assessment.

Multidimensional pain assessment scales, such as the McGill Pain Questionnaire, were not used due to the fact that they are very time consuming, they require large sample sizes, and have been tested only for English speaking populations. In addition, even if the person is fluent in English, the meaning of the words may differ from one culture to the other (Chapman et al., 1985; Gerschman et al., 1987). Considering that a multicultural society was being studied, the most simple scales with the most simple wordings were chosen to minimize the cultural interpretation differences. Additionally, the neuropsychological tests I used demand energy and attention, and therefore long pain inventories might have induced fatigue prior to the beginning of the test, which could have affected the neuropsychological test results (Stuss et al., 1985, 1987; Goldberg et al., 1996).
c) Improvement criteria

Thirty percent reduction as a percentage of the baseline assessment in pain at rest (100 mm Visual Analogue Scale) was used as the criteria for improvement. Visual analog scales provided a valid, reliable means of assessment of patient perceptions of treatment outcome (McCreary et al., 1992; Price et al., 1994). This cut off point was used, because in one study approximately seventy percent of patients had a 30% or more reduction (VAS ratings) in pain at rest (Dao et al., 1994); which was consistent with the available literature on TMD treatment success rate (60 - 90%) (Greene & Laskin, 1983). In addition, this level of improvement in VAS scores (30%) could not be explained only by the role of chance. The ICC for VAS was very high (0.91) in a range of ± 5 mm for acute or experimental pain (Price et al, 1994). In addition, the variation in pain intensity recall post-treatment versus baseline measures was reported not to exceed 23%. Therefore, an improvement of 30% or greater is larger than both the variation in the VAS measurements and pain memory.

The measurements were made at baseline and at a six-month follow-up. Repeated measurements were avoided, because this was shown to influence the degree of correlation between two different scales during the course of treatment (Carlsson, 1983). This phenomenon happened, because the patient might have learned the use of the scale.
d) Description of the populations studied

Patients with TMD who met our improvement criteria were included in Group I (responding or rTMD). While those who did not improve were included in Group II (non-responding or nrTMD). Patients with irritable bowel syndrome (IBS) were part of Group III. Finally, a non-pain population was also included and became part of Group IV.

i) Temporomandibular disorder group (Groups I and II): inclusion criteria

Patients who participated in this study were those seeking treatment for temporomandibular disorders (TMD) at the Craniofacial Pain Research Unit at the Mount Sinai Hospital, and the Facial Pain Clinic at the University of Toronto Faculty of Dentistry. The cases were newly diagnosed to avoid selection biases and fluctuation of signs and symptoms of TMD. Usually, in newly diagnosed cases, the diagnosis is more uniform, and the patient recall is better which prevents recall bias, and the examiner is more certain that the recall preceded the diagnosis (Schleselman, 1982).

The target population, based on the clinical examination, were (1) women between the age of 15 and 45 years old; and (2) chief complaint of frequent pain (at least 4 times/week) in the masticatory muscles and/or temporomandibular joint area; (3) tenderness to palpation of at least 3 sites in the masticatory muscles and/or the
temporomandibular joint area; and/or (4) limitation in normal mandibular movement of less than 40 mm (Dworkin & LeResche, 1992; Dao et al., 1994; Goldberg et al., 1996).

ii) TMD exclusion criteria (Groups I and II):

Patients were excluded based on the medical history from the study if they had the following conditions. First, patients diagnosed with acute muscle spasm, myositis, contracture, polyarthritis (osteoarthritis and osteoarthrosis), and acute traumatic injury prior to the use of the RDC/TMD (Dworkin & LeResche, 1992). Second, patients who had already been treated with occlusal splints; who had any previous treatment for TMD; who were wearing complete prosthesis or removable partial prosthesis with distal extensions. Third, patients with medical and/or dental emergency, metabolic diseases (e.g., diabetes, hyperthyroidism), neurologic disorders (e.g., dyskinesia, trigeminal neuralgia), vascular disease (e.g., migraine, hypertension), neoplasia, history of psychiatric disorders, history of drug abuse, motor vehicle accident (M.V.A.), currently receiving medication or other treatments (e.g. acupuncture, physical therapy), and history of allergy to acrylic (Dao et al., 1994; Goldberg et al., 1996). Fourth, patients with TMJ pain without involvement of the masticatory muscles were also excluded. Regarding pain medication, only those with effects on the CNS (muscle relaxants, anti-convulsants, and anti-depressants) were excluded. Patient taking pain medication such as analgesics and anti-inflammatories were included, but a wash-out period of 3 days prior to neuropsychological testing was required. Finally, patients with reported major visual,
auditory, and motor impairments were also be excluded, considering that this might have affected the neuropsychological test performance (Goldberg et al., 1996).

iii) Irritable bowel syndrome group (Group III): inclusion criteria

The IBS patients were diagnosed by the treating clinician. The IBS patients were women only, previously treated for the condition, between the ages of 15 and 45 years, and referred by gastroenterologists from major gastroenterology outpatient clinics in general hospitals. Patients were also recruited by newspaper adds. When the patient was recruited from newspaper adds; the treating clinician was contacted, with previous authorization of the patient, in order to confirm the diagnosis that the patient had been previously treated for the condition. The IBS patients was diagnosed by the specialist, from both specialist referrals and newspaper adds. They were selected by the examiner based on the medical history following the Rome guidelines respectively (Drossman, Li & Toner et al., 1995). This category included persons who experienced abdominal pain more than six times in the prior year, in combination with two or more of the following symptoms (referred to as the Manning symptom criteria): 1) pain that was often relieved by defecation (more than 25 percent of the time); 2) looser stools often when pain began; 3) more frequent stools often when pain began; 4) abdominal distension often; 5) a feeling of incomplete evacuation often; and 6) mucus per rectum. Based on the available literature, a cutoff score of two or more criteria was used to identify symptoms compatible with IBS, as this was considered optimal for epidemiologic studies.
iv) Irritable bowel syndrome (Group III): exclusion criteria

The irritable bowel syndrome group (IBS) was selected by the examiner based on a phone interview. The exclusion criteria was the same described for the TMD group with the exception of the dentistry-related variables examined which were listed again to prevent confusion of which were used or not. First, those treated with occlusal splints; those who had any previous treatment for TMD. Second, those who were wearing complete prosthesis or removable partial prosthesis with distal extensions. Third, patients with medical and/or dental emergency. Fourth, those who had history of allergy to acrylic (Dao et al., 1994; Goldberg et al., 1996). Finally, patients with TMJ pain without involvement of the masticatory muscles.

v) Non-pain population (Group IV): rationale

Although published external control values could have been used, neuropsychological tests may be sensitive to different examiners, different test sites, and different test protocols (Stuss et al., 1985). Therefore, an internal control group was required for comparison; however, comparisons between our non-pain group scores with external norms were made, as well to determine how my test population compared to other groups tested previously. The primary goal of the comparisons with non-pain population was to evaluate if the pain level, measured by 100 mm VAS, at rest post-
treatment of the responding TMD group (rTMD) was reduced to ‘normal’ levels. This was necessary, because when considering treatment efficacy, it is important not only to assess improvements in the responding TMD patients over baseline scores or versus non-responding TMD group values, but also in comparison with non-pain individuals (Jacobson et al., 1984). Finally, this assessment was used also to evaluate if the scores of the responding TMD group neuropsychological test scores were more similar to the non-pain group than with the non-responding TMD.

vi) Non-pain population (Group IV): inclusion criteria

The non-pain population included were women only, previously treated for the condition, and between the ages of 15 and 45 years (age and sex matched at baseline). The participants were volunteers selected by the examiner based on a phone interview. The non-pain group was selected from the Faculty of Dentistry staff and students, the Mount Sinai Hospital staff, and advertisements posted around the University of Toronto Main Campus.

vii) Non-pain population (Group IV): exclusion criteria

The exclusion criteria was the same described for the IBS group, with the additional criteria that they must have had no sleep disorders and with no previous
treatment for chronic pain conditions, such as irritable bowel syndrome and temporomandibular disorders.

e) Instruments

The instruments used as baseline measures were described as follows: i) neuropsychological assessment for tests measuring short-term memory, short-term memory under interference, verbal and auditory memory, and attention (Reaction Time Tests, California Verbal Learning Test, Brown-Peterson Consonant Trigram Auditory Memory Task), ii) psychosocial assessment for sleep and depression only (Beck Depression Inventory and Sleep Assessment Questionnaire, respectively), iii) clinical assessment for signs and symptoms of TMD (palpation of the TMJ facial, TMJ auditory meatus, TMJ sounds, masseter, temporalis, sternocleidomastoid, medial pterygoid, lateral pterygoid, and insertion of the temporalis), and iv) occlusal examination (maximum mouth opening, overbite, overjet, exacerbation after examination, percussion sensitivity and caries). The test descriptions and their rationale will be described below. In this study, the operational definition of neuropsychological 'deficits' describes significant differences between and among groups, both in the chronic pain patients as well as in the non-pain population and may therefore also be taken to mean low-normal range (and statistically significant). Thus this term will not necessarily mean that any one group’s neuropsychological scores are two standard deviations, plus or minus, outside the normal range, although this may be the case in some instances.
i) Neuropsychological assessment

The tests used in the study were based on a previous study (Goldberg et al., 1996) which employed neuropsychological tests for comparison of post-traumatic TMD (pTMD) population versus and idiopathic (iTMD) one. The tests employed in this study were the ones which have shown significant differences between the two groups. Simple Reaction Time test (SRT), Multiple Choice Reaction Time test (MCRT), Multiple Choice Reaction Time test with Conflict (MCRTCF), and Multiple Choice Reaction time test with Constraint (MCRTCT) - scores (0 - ∞) have been used for individuals with mild head injury who develop difficulties in information processing and showed an average significant difference of $P < .05$ between the two groups. The California Verbal Learning Test - Correct Responses (CVLT-TCR, scores 0 - 80), the California Verbal Learning Test - Cluster (CVLT-TC, scores 0 - 60), the California Verbal Learning Test - Perseveration (CVLT-P, scores 0 - 40), and the California Verbal Learning Test - Intrusion (CVLT-I, scores - 10) were designed to detect deficits in encoding verbal information, as well as transferring the information to short- and long-term memory, and these tests showed a significant difference in the immediate recall ($P < .05$) in the learning phase. In addition, highly significant differences ($P < .001$) were also found in the Brown-Peterson Consonant Trigram Auditory Memory Task (CCC, scores 0 - 45) which evaluates memory under interference. These tests were also showed in study of the
neuropsychological profile of patients with good recovery after closed head injury (Stuss et al., 1985, 1989a) to have reproducibility of the results.

A personal computer controlled stimuli for the Reaction Time tests. Stimuli were white or colored on a constant background of dark grey and displayed on a 35 cm color monitor situated 1.5 m from the subject. The approximate size for each stimulus was 5 cm square. The mean interstimulus interval was five seconds with a total range of mean stimuli interval of two seconds. Subjects pressed a button in their preferred hand for the Simple Reaction Time (SRT) test; for the Multiple Choice Reaction Time (MCRT) tests, button responses were required by both hands in the MCRT tests and with only one hand in SRT tests. (Stuss et al.,1989b).

In the simple reaction time test (SRT), the subject was asked to press a button in her preferred hand, as quickly as possible, in response to the presentation of a stimulus. The stimulus was randomly selected from among four designs (a circle, square, triangle or cross) and was constant throughout the test. All stimuli were white outlines without shading. Five practice trials were followed by 50 test trials. The dependent measure for this test was the mean reaction time in milliseconds (Stuss et al., 1989b).

In the multiple choice reaction time tests (MCRTs), three MCRT tests - easy, with conflict, and with contraint - were administered. The stimuli were either a target or nontarget. The stimuli are randomly presented. The target stimulus had a 25% probability of presentation and was randomly selected prior to test onset. The subject pressed the button in the preferred hand in response to a target and the button in the other hand in response to a nontarget. In each case 10 practice trials were followed by 100 test trials.
For each MCRT test, only the target correct response times were analysed for sake of simplicity.

In the plain MCRT, one of the four white geometric shapes (a circle, square, triangle and cross) was randomly selected as the target, the remaining three being nontargets.

In the Multiple Choice Reaction Time Test with Conflict, the stimuli had three different components (shape, colour, and line orientation within the shape), each of which could appear in one of four possible states. The shape could be a circle, square, triangle or cross. The colour could be red, blue, green or yellow. The line orientation could be vertical, horizontal, backward slanting (\) or forward slanting (/). The target possessed a randomly selected combination of these states, that is, a blue circle filled with vertical line. Non-targets were stimuli that did not possess all three of the states belonging to the target. For example, relative to the aforementioned target, the following would be nontargets: a red circle with vertical lines; a blue square with horizontal lines; a yellow triangle with backward slanting lines. The probability of target stimulus as well as the probability that each property (e.g. color, line, shape, and orientation) of the non-target stimulus is identical to that of the target stimulus is 25%.

In the Multiple Choice Reaction Time Test with Constraint, the stimuli in this case were characterized by three components as in the MCRT with Conflict. However, no state specific to the target could ever appear in a nontarget. For example, if the target was a red circle with vertical lines, no nontarget would be red, be a circle or possess vertical lines. Subjects were informed of these constraints but were not instructed to focus on any
one state. Hence the stimuli appeared to be as complex as in the MCRT with Conflict but most of the information provided was redundant. If the subject focused on only one stimulus state at a time, the tested reverted to the same level of difficulty as the plain MCRT (Stuss et al., 1989b). The probabilities are identical (i.e. 25%) to the ones described for the MCRTCF.

The standard norms for the reaction time tests were based on a study by Stuss et al. (1989a) which used 20 controls are the following (mean, SD): a) Simple Reaction Time test (243,67), b) Multiple Choice Reaction Time test (433, 54), c) Multiple Choice Reaction Time test with conflict (512, 78), and d) Multiple Choice Reaction Time test with constraint (440, 68).

The CVLT was designed to evaluate multiple cognitive parameters (the process of verbal learning and the amount of material acquired and retained) using an everyday verbal memory task (Delis et al., 1987). The subject was presented with a “Monday” list of 16 items (four each, in the categories of clothing, spices/herbs, tools, and fruits) over five trials. Several dimensions of performance can be evaluated, including semantic and serial learning strategies, retention of information over time, and free versus cued recall versus recognition memory. Normative data are available across the age range of 17 to 80. The CVLT has good split-half reliability (0.92). Test-retest values are somewhat lower (e.g. List A total recall = .59); however, the retest interval was 1 year. The CVLT correlates significantly with various Wechsler Memory Scale (WMS) subtests (Delis et al., 1988). The test manual summarizes CVLT data on a variety of clinical populations,
including alcoholism, Parkinson’s disease, multiple sclerosis, Huntington’s disease, and Alzheimer’s disease.

Additional research further supports the construct and criterion-related validity of the CVLT. Delis et al., (1988) analyzed 286 normal subjects (105 men and 181 women; mean age = 60.20 years, range = 19-91) and 113 neurological patients (72 men and 41 women; mean age 51.7 years, range = 9-20). The patient group consisted of 55 patients with multiple sclerosis, 8 with Huntington’s disease, 24 with chronic alcoholism and 26 with Parkinson’s disease. Using factor analysis, they reported a good factor structure representing several underlying memory processing components, including general verbal learning, response discrimination, proactive effect, and serial position effect, among others for both patient and controls.

Additionally, compendia on the CVLT (Mapou & Spector, 1995; Sbordone & Long, 1996) reported data comparing the performance of severe traumatic brain injury (TBI) patients with neurologically normal male adults. These subject groups differed on the CVLT in both level and process of verbal learning (i.e., TBI subjects as a group did not typically use semantic grouping strategies). In addition, using the CVLT scores, it was possible the correct classification of over 76% of cases of Huntington’s disease (HD), AD, and Parkinson’s disease (PD) on the basis of CVLT performance. Significant correlations between CVLT and Verbal Selective Reminding Test (VSRT) scores were reported.

Published norms, females only with age range 17 to 44 years, for the mean CVLT correct responses ranged from 62 to 64 (T50). However, standard deviation for the
CVLT correct responses, mean scores and standard deviations for the CVLT sub-scales (clusters, perseverations, and intrusions) were not available (Delis et al., 1987).

For studying short-term retention, a popular method has been this distractor technique which is also called the “Peterson task”, the “Peterson and Peterson procedure” and other variations on the Peterson name, or it may be referred to as “consonant trigrams.” The purpose of the distractor task was to prevent rehearsal of material being held for short-term retention testing (Peterson & Peterson, 1959). The test measures auditory short-term memory of three consonants under interference conditions. A trigram (three letters) was delivered to the subject verbally at a rate of 1 letter/second followed immediately by a 3-digit random number. The subject was asked to count backwards outloud by threes for random interval delays of 9, 18, and 36 seconds until signalled to stop counting. The subject was then asked to recall the trigram. Five trials were given for each delay with intertrial delays of 2-5 seconds. The select delays exceeded the normal delays of 3, 9, 18 seconds to minimize any ceiling effect. Dependent measures were the total number of correct letters recalled at each of the three delay intervals. The maximum score at each delay interval was 3 and the total score was 45 (Stuss et al., 1987). For example, if the test item was three consonants, the examiner said, “V J R, 386” and the subject began counting until stopped at the end of the pre-designated time interval to recall the item.

Stuss and colleagues, 1982, have used the Peterson-Peterson Consonant Trigram test procedure (Peterson & Peterson, 1959) to evaluate short-term memory in frontal-loobotomized schizophrenics. In this procedure, subjects were provided with three
consonants and then engaged in an interfering activity (e.g., counting) for 3, 9, and 18 seconds. Consonant Trigrams was the only test out of several measures of learning and memory that was sensitive to orbitofrontal lobotomy. With this technique, normal subjects had perfect recall with no distraction delay. They recalled about 80% of the letters correctly with a distraction duration of 3 sec, and a 70% to 80% correct recall with 9 sec delays (Stuss et al., 1987). Longer durations produced a wider range of normal performances: From 50% to 80% with delays of 18 sec, and around 67% with the delays of 18 sec, and around 67% when the delay is as long as 36 sec. Giving five trials for three consonants each for a total of 15 possible correct responses at each delay interval, Stuss and coworkers, 1989a, report standard deviations typically within the 1.9 to 3.2 range. Subsequently, they analysed recovery patterns on the Trail Making Test and the Consonant Trigrams procedure in patients with mild and severe TBI. Both tests significantly discriminated control subjects from severe TBI patients, whereas only Consonant Trigrams discriminated mild TBI patients from controls. Stuss et al., 1987, 1988) provided normative data for a 9-, 18-, and 36-second version of the test.

Differences in sex, age- from late teenagerhood to 69 years- or educational levels (high school completion or less versus more than high school) were not statistically significant. However, by inspection women showed a tendency to better recall than men. Individuals with more than a high school education had slightly higher scores on average, and the older subject groups did a little less well than younger ones. Small but significant practical effects were documented (Stuss, Stethem and Poirier, 1987).
The published norms for the Brown-Peterson Consonant Trigram Auditory Memory Task (CCC) were the following: i) 20 to 40 years old, mean = 43.4 and SD = 2.6; and ii) females only, mean = 35.0 and SD = 2.2 (Stuss et al., 1987, 1988).

ii) Psychosocial assessment

The psychosocial assessment was made only made for fatigue, energy level, depression and sleep as possible predictors of TMD treatment outcome (Gale & Funch, 1984; Gerschman et al., 1987; Fricton & Olsen, 1996). These variables were chosen, because they have shown to affect the test results of both the reaction time tests, and the California Verbal Learning Test, and the Brown-Peterson Consonant Trigram Auditory Memory Task (Stuss et al., 1987, 1989a, 1989b; Stuss & Levine, 1995; Goldberg et al., 1996). In addition, they were not only analyzed as predictors but also as confounders for all neuropsychological tests described.

In order to assess depression, the long form of the Beck Depression Inventory (scores 0-63) was used. The purpose of this test is to screen for depression by self-report statements. The long form is more frequently employed and has already been validated against the DSM-III (sensitivity=0.83, specificity=0.89, PPV=0.65, NPV=0.95)(Rugh, 1987).

The patient checked 21 four-choice statements presented on a single age for the choice or choices most appropriate to him or her. The statements refer to the following
areas: 1) sadness, 2) pessimism/discouragement, 3) sense of failure, 4) dissatisfaction, 5) guilt, 6) expectation of punishment, 7) self-dislike, 8) self-accusation, 9) suicidal ideation, 10) crying, 11) irritability, 12) social withdrawal, 13) indecisiveness, 14) unattractiveness, 15) work inhibition, 16) insomnia, 17) fatigability, 18) loss of appetite, 19) weight loss, 20) somatic preoccupation, and 21) loss of libido (Beck, 1970).

In the administration of the test, the examiner says to the patient: “This questionnaire consists of twenty-one groups of statements. After reading each group of statements carefully, circle the number - zero, one, two, or three - next to the one statement in each group that best describes the way you’ve been feeling in the past week, including today. If several statements within a group seem to apply equally well, circle each one. If the patient indicates his or her choice by responding with a number, read back the statement corresponding to the number given by the patient, to clarify exactly which statement the examinee has selected. When the patient responds, “The first statement,” he or she may mean (0) or (1). After it is apparent that the patient understands the numbering system, the numerical answer should be sufficient to indicate his choice. The BDI may be given to the patient for self-administration or group administration, but it should be verified that the patient understands the purpose and the answering method for the test as outline above. The approximate time for administration is between 5 to 10 minutes. The total score is obtained by adding the highest score circled for each of the 21 items. The maximum score is 63. Item 19 (weight loss) was designed to assess anorexic symptoms. If the patient responds affirmatively to the supplementary question “Are you
trying to lose weight by eating less?" the score on that group is not added to the score. Be sure to read all the statements in each group before making your choice (Beck, 1970)."

At this point, hand a copy of the questionnaire to the patient and say: "Here is copy for you, so that you can follow along as I read." Read the entire group of statements in the first category (do not read the numbers appearing before the statements), then say: "Now, which one of the statements best describes the way you have been feeling in the past week, including today?"

The BDI is just one of a score of depression scales (e.g., Hamilton, 1967; Radloff, 1977) developed to detect depression in routine screening or research. It was selected because of its simplicity of administration, scoring, and interpretation. Since the items are very similar to many MMPI items, it need not be given if the MMPI is administered. Moreover, depression has been recognized as a multidimensional disorder. Bolon and Barling (1980), for example, extracted three factors (ideational depression, physiological depression, behavioral depression). Others derived five factors from the Zung scale; and the MMPI delivers several subscales (pure depression, subjective depression, psychomotor retardation, physical malfunctioning, mental dullness, brooding), in addition to the D-scale (Scale 2), that allow differential diagnostic considerations which screening inventories cannot provide because of their brevity (Sbordone & Long, 1996).

The test-retest reliability with 38 patients was above 0.9 and tended to follow the trend for each patient on depth of depression (Beck, 1970). Spearman-Brown reliability was 0.93, and internal consistency for test items 0.86 (Reynolds & Gould, 1981). Concurrent validity coefficients with Lubin's Depression Adjective Checklist were 0.38-
0.50 for psychiatric patients and 0.66 in controls, with the Zung Self-Rating Depression scale, 0.79 in psychiatric patients and 0.54 in college students with the MMPI D-Scale, 0.75; and 0.78 and 0.82 in psychiatric patients (Mapou & Spector, 1995). Beck (1970) also reported correlations of 0.66 between the BDI and psychiatric ratings of university students. The test also overlapped with the Beck anxiety checklist (0.60) and the Maudsley obsessive-compulsive index (0.49 in nonclinical populations)(Sbordone & Long, 1996). It had only a modest negative correlation (-0.41) with Rotter’s (1966) Internalizing-Externalizing Scale and with Duttweiler’s (1984) Internal Control Index (-0.37), suggesting less depression in persons with internal control (Meyers & Wong, 1988).

There is no arbitrary score that can be used for all purposes to classify different degrees of depression. However, some studies suggested guidelines to interpret the long form. The interpretation guidelines for the long form of the Beck Depression Inventory are the following: a) normal range (0-9), b) minimal depression (10-15)(cutoff=10.9, SD=8.1), c) mild-moderate depression (16-19)(cutoff=18.7, SD=10.2), d) moderate-to-severe depression (20-29)(cutoff=25.4, SD=9.6), and e) severe depression (30-63)(cutoff=30.0, SD=10.4) (Beck, 1970).

Few specific instruments have been developed for evaluation of fatigue (Yoshitake, 1978); however, usually they are time consuming and difficult to interpret which can interfere with neuropsychological tests administration afterwards. In addition, they have not been validated for use in TMD populations. In a compendium of neuropsychological tests, no specific instruments for measuring both fatigue and energy
levels have been developed (Kolb & Whishaw, 1989; Spreen & Strauss, 1991). They are usually sub-scales of personality inventories such as the MMPI. Given this, and to avoid unduly fatiguing test subjects, both psychosocial variables were assessed simply by visual analog scales (100 mm VAS).

Regarding sleep assessment, the use of a reliable and validated sleep questionnaire is very important, because although important findings were obtained, previous studies that identified sleep as a predictor of TMD treatment outcome had not used such an instrument (Fricton & Olsen, 1996). In order to assess sleep, the University of Toronto Sleep Assessment Questionnaire (SAQ) was used as this has been validated (Cesta, Moldofsky & Sammut, 1996). The SAQ is a 19 item self-administered scale. Regarding reliability, of the 77 patients who completed the first copy of SAQ, 68 returned their second copy (88%). The intra-class correlation coefficient was 0.97, a value characterized as almost perfect by Landis & Koch (1977). The five factors that were identified within the SAQ were labeled: (i) non-restorative sleep, (ii) sleep schedule disorder, (iii) disturbed sleep, (iv) sleep apnea, and (v) hypersomnolence. The coefficient alpha did not increase when individual items were removed from each of the above factors; therefore, the questions were homogeneous, and the initial version of the SAQ was not modified. SAQ has favorable criterion-related validity when correlated (coefficient alpha) with non-restorative sleep (R=0.67, p<0.0001), disturbed sleep (R=0.63, p<0.0001); and hypersomnolence factor (R=0.49, p<0.0001)(Cesta, Moldofsky & Sammut, 1996).

Normative data have been published (Cesta, Moldofsky & Sammut, 1996) for 289 patients and 30 controls. The primary diagnosis of the patients were sleep apnea (SA),
periodic leg movements (PLMS), snoring (SN), and normals. Receiver operator curves (ROC) were constructed using sensitivity and specificity of the total SAQ score (ranging from 0 to 68). The following cutoff points were described: i) 17 (sensitivity = 0.71, specificity = 0.83), ii) 16 (sensitivity = 0.73, specificity = 0.80), and iii) 14 (sensitivity = 0.80, specificity = 0.73). Our choice of cutoff point will be 16, because when our objective is to differentiate between two patient populations, we should also choose the point with the highest sensitivity. The reported norms are for the SAQ total score are (mean, SD): i) patient population (26.0, 8.6), and ii) control population (10.8, 5.7).

iii) Clinical examination

The clinical examination procedures and TMD measurement criteria used were based on the RDC/TMD (Widmer, 1992a). This procedure was done to allow comparison with other studies which evaluated similar neuropsychological tests and signs and symptoms for TMD populations.

Traditional signs and symptoms usually recorded by dentist when examining TMD patients have not been shown to be good predictors of treatment outcome (Dworkin et al. 1989, 1991); however, this isolated finding still needs confirmation. Palpation of the temporomandibular joint and masticatory muscles were included due to the fact that these variables have been included in almost every study of temporomandibular disorders and are part of the standard examination procedure for TMD (Dworkin & LeResche, 1992).
In order to determine whether there might be differences in the musculoligamentous components and other clinical aspects of our TMD population in addition to the neuropsychologic differences, subjects entering this investigation underwent complete extraoral and intraoral clinical examinations. The extraoral examination included palpation of the masseter, temporalis and sternocleidomastoid muscles, as well as palpation of the TMJ itself by one examiner. In addition, the medial pterygoid muscle, the lateral pterygoid muscle (or region), and the insertion of the temporalis muscle at the coronoid process were palpated (Goldberg et al., 1996).

All scores assigned were based on the patient's responses when the sites were palpated (i.e., evoked pain reaction). A scale of 0 to III was established (0 meaning no pain response). A grade I pain response was considered a mild observance that discomfort was present in that the patient had to be asked whether pain was felt. A grade II was assigned when changes in facial expression connoting a pain reaction (or verbal pain reaction) were produced (i.e., the patient did not have to be asked). A grade of III was scored when definite avoidance to palpation was observed or when normal palpation force was abated before the patient reacted too violently. In order to increase our reliability and for analytical purposes, scores were recoded into two groups: grade 0-I was considered as a negative pain reaction score; and grade II-III was considered as a positive reaction score.

Temporomandibular joint sounds (clicking, popping or crepitation) also had scores ranging from 0 to III; however, their meanings differed from muscle and joint palpation because we were not assessing pain. Grade 0 meant no audible or palpable joint
sounds both by the clinician and the patient when the patient opens and closes his/her mouth. Grade I meant no audible or palpable joint sounds by the clinician, but the patient identifies gentle sounds. Grade II meant palpable joint sounds by the clinician and confirmed by the patient. Grade III meant audible sounds detected by both the clinician and patient and confirmed during palpation. The scoring also followed the same procedures described for masticatory muscle and TMJ pain (i.e. Grade 0-I, negative test result, and Grade II-III, positive test result).

The extra-oral muscle palpation reliability (kappa = 0.47 to 0.65) indicates acceptable agreement (0.6 to 1.0 is good to perfect agreement according to Landis & Koch, 1977) with calibrated examiners in both symptomatic and asymptomatic populations. The intra-oral muscle reliability is lower than the extra-oral (kappa = 0.27 to 0.61), but the TMJ palpation achieves similar standards (kappa = 0.47 to 0.52). The auscultation of the TMJ (kappa = 0.62 by direct palpation and 0.61 for stethoscope) sounds also reached acceptable levels (Widmer, 1992b).

Reliability score is a consequence of intra-examiner and inter-examiner reliability, stability of the phenomenon being measured over time, and reliability of patient report to pain. Since muscle and TMJ palpation responses can vary from one exam to another during the same day or from one day to the next, the difficulty with obtaining high reliability scores is obvious (Widmer, 1992b). In order to overcome these problems, a single composite score for muscle and joint palpation has been recommended for higher reliability scores (ICC = 0.87 to 0.91). In our study, we also included a muscle and joint combined score, which were a combination of the individual scores, and the cutoff points
also followed a similar combination of the cutoff points of each individual examination (Goldberg et al., 1996). Considering that six masticatory muscles were palpated, in our combined muscle score, scores from 0 through 6 meant a negative result, and scores from 7 through 18 meant a positive test result. In our combined joint score, scores ranging from 0 through 3 were considered a negative result, while those ranging from 4 through 9 meant a positive test result.

The extra-oral palpation was done with the patient in resting position with the mouth closed: origin and insertion of the masseter muscle as well as its central portion; anterior, medial and posterior portions of the temporalis muscle, origin and insertion of the sternocleidomastoid muscle as well as its central portion.

Examining the muscles and joint capsules or tenderness requires that the examiner presses on a specific site using the fingertips of the index and third finger or the spade-like pad of the distal phalanx of the index finger only with standardized pressure, as follows: palpations were done with 2 lbs of pressure for extra-oral muscles, 1 lb of pressure the joints and intra-oral muscles. Palpation of the muscles while using the opposite hand to brace the head was performed to provide stability. The subject's mandible should be in a resting position, without the teeth touching. Palpate while muscles are in a passive state. As needed, have the subject lightly clench and relax to identify and to insure palpation of the correct muscle site. ("I'm going to press on some muscles. I would like you to clench your teeth together gently and then relax and have your teeth slightly apart from each other.") First locate the site of palpation using the landmarks described and then press. Because the site of maximum tenderness may very
from subject to subject and is localized, it is important to press in multiple areas in the region specified to determine if tenderness exists. Before beginning the palpations, say: “In the next part of the exam, we’d like you to record whether you feel pain or pressure when I palpate or press on certain parts of your head and face.” Ask the subject to determine if the palpation hurts (painful) or if he/she just feels pressure. If it hurts, ask the subject to indicate if the pain is mild, moderate, or severe. Record any equivocal response or the report of pressure only as “No Pain.”

Our extra-oral examination included the following masticatory muscles and areas related to the temporomandibular joint:

a) Palpation of the TMJ lateral pole: place your finger just anterior to the tragus of the ear and over the subject’s TMJ. Ask the subject to open slightly until you feel the lateral pole of the condyle translated forward. Use 1 lb pressure on the side that is being palpated, supporting the head with the opposite hand.

b) Posterior attachment of the TMJ: this site can be palpated inside the external auditory meatus. Place tips of the right little finger into the subject’s left external meatus and the tip of the left little finger into the subject’s right external meatus. Point the fingertips toward the examiner and ask subject to slightly open the mouth (or wide open if necessary) to make sure the joint movement is felt with the fingertips. Place firm pressure on the right side and then the left side while the subject’s teeth are completely together.
c) TMJ sounds: Subjects indicate the presence or absence of sounds; if present, the examiners will score the type of sound observed. Place the left index finger over the subject’s right TMJ and the right index finger over the subject’s left TMJ (preauricular area). The pad of the right finger is placed anterior to the tragus of the ear. Ask the subject to slowly open as wide as possible, even if it causes pain. Each closure should bring the teeth completely together in maximum intercuspation. Ask the subject: “While I have my fingers over your joint, I would like you to slowly open as wide as you can and then slowly close until your teeth are completely together.” Ask the subject to open and close three times. Record the action/sound that the joint produces on opening or closing as detected by palpation and as defined: 0 = none, 1 = Click. A distinct sound, of brief and very limited duration, with a clear beginning and end, which usually sounds like a “click”. Circle this item only if the click is reproducible on two of three openings/closings, and 2 = patient with coarse and fine crepitus will be excluded.

d) Masseter muscle: origin, ask the subject to first clench then relax and observe masseter for location. Palpate the origin of the muscle beginning in the area 1 cm immediately in front of the TMJ and immediately below the zygomatic arch, and palpate anteriorly to the border of the muscle. Body of masseter. Start just below the zygomatic process at the anterior border of the muscle. Palpate from here down and back to the angle of the mandible across a surface area about two fingers wide. Insertion of the masseter. Palpate the area 1 cm superior and anterior to the angle of the mandible.

e) Temporalis muscle: for the temporalis posterior, palpate posterior fibers behind the ears to directly above the ears. Ask the subject to clench and then relax to help
identify muscle. Walk finger toward the subject’s face (medially) to the anterior border of the ear. Temporalis middle: palpate fibers in the depression about 2 cm lateral to the lateral border of the eyebrow. Temporalis anterior: palpate fiber over the infratemporal fossa, immediately above the zygomatic process. Ask the subject to clench and relax to help identify muscle.

The intra-oral palpation with the patient with the mouth opened: medial pterygoid area, lateral pterygoid area, coronoid insertion of the temporalis.

f) Medial pterygoid area: ask the patient to open his mouth as wide as possible and palpate the index finger with gentle pressure in the area closer the angle of the mandible under the tongue and observe any jerking movements or ask the patient if there is any presence of pain at all.

g) Lateral pterygoid area: before palpating, make sure the fingernail of the index finger is trimmed to avoid false positives. Ask the subject to open the mouth and move the jaw to the side that is being examined. (‘Move your jaw toward this hand.’) Place the index finger on the lateral side of the alveolar ridge above the right maxillary molars. Move the finger distally, upward, and medial to palpate. If the index finger is too large, use the little finger (5th digit).

h) Insertion or tendon of temporalis: after completing the lateral pterygoid, rotate your index finger laterally near the coronoid process, ask the subject to open slightly, and
move your index finger up the anterior ridge of the coronoid process. Palpate on the most superior aspect of the process. If it is difficult to determine in some subjects if they are feeling pain in the lateral pterygoid or the tendon of the temporalis, rotate and palpate with the index finger medially and then laterally. If there is still difficult, the lateral pterygoid is usually the more tender of the two.

Intraoral examination included a complete dental examination to rule out pain from dentoalveolar etiology (Goldberg et al., 1996). Similar to palpation, other signs which are usually detected and recorded by the dentist during clinical examination were also included in our study. These signs were mandibular range of motion (maximum unassisted mandibular opening, overjet, exacerbation after examination, caries, and percussion sensitivity). In addition pain intensity at rest and on chewing pre- and post-treatment were also assessed. Considering reliability of the measurements, maximum unassisted opening reached a very high level of agreement with trained examiners (ICC = 0.96) and with untrained examiners (ICC = 0.90) (Widmer, 1992b).

The occlusal variables (intraoral examination) to be performed by the dentist were the following:

a) Maximum unassisted mandibular opening: with a 100 mm ruler from the incisal border of the upper and lower central incisors minus the overbite. Ask the subject to place the mandible in a comfortable postion. (“Place your mouth in a comfortable
position'.) Then ask the subject to open the mouth as wide as possible, even if he/she feels pain. ("I would like you to open your mouth as wide as you can, even if it's a little uncomfortable.") Place the edge of the millimeter ruler at the incisal edge of the maxillary central incisor that is the most vertically oriented and measure vertically to the labioincisal edge of the opposing mandibular incisor; record this measurement.

b) Overbite: with a 100 mm ruler from the incisal border of lower central incisor to the point of overpassing of the incisal edge of the upper central incisor divided by the tooth length. Ask the patient to close the teeth completely together. With a pen or fingernail, mark the line where the incisal edge of the same maxillary central incisor used before for measurements overlaps the mandibular incisor. Measure the distance from the mandibular incisal edge to the marked line and record the measurement.

c) Overjet: with a 100 mm ruler from the incisal border of the upper central incisor to the buccal surface of the lower incisor divided by the tooth length.

d) Caries: presence or absence of carious lesions with a probe by the dentist (yes or no).

e) Percussion sensitivity: presence or absence of percussion sensitivity with the handle of a mirror by the dentist (yes or no).
Regarding pain intensity, the following variables were included:

f) Pain intensity at rest and pain on chewing (pre- and post-treatment) was assessed by self-completing visual analog scales (100 mm VAS).

g) Exacerbation of orofacial pain after clinical examination: patient was asked by the dentist and must answer either yes or no.

f) Confounders

Neuropsychological tests are extremely sensitive to many different confounders. Variables such as age, gender, language, chemical dependency, history of trauma, neurological disorders, psychological status have been shown to influence cognitive tests (Stuss & Levine, 1995; Goldberg et al., 1996). These variables were controlled in the design stage by restriction in the inclusion/exclusion criteria and during the medical history. Time of the day and the site in which the test is performed may also influence the test results due to diurnal pain fluctuation and external distraction and were controlled in the protocol. All tests were scheduled between 10 am and 4 pm; because, during this time period, the pain fluctuation is minimal (less than 10mm on a 100mm VAS) (Folkard, 1976).

Other confounders that have been reported to influence specifically the neuropsychological tests to be used in our study that still remained after the design stage
are educational level (substitute for IQ), employment, and income level (Stuss et al. 1985; Stuss & Levine, 1995). Age will be controlled at the designed stage and reassessed at follow-up in order to verify if the different groups are statistically different. Their influence on the results were controlled in the analysis stage and assessed both in the bivariate and in the logistic regression analyses (Hennekens & Buring, 1987; Norusis, 1991, 1992). They were not controlled in the design stage to prevent excessive reduction in our sample size. They were assessed by self-reported questionnaires.

Finally, other variables, despite not being confounders for the neuropsychological tests, that may bias our results are length of treatment, pain duration, number of treatments, treating clinician and comorbidity with irritable bowel syndrome. They were assessed by reviewing the patient’s medical and orofacial pain history. Finally, the depression, sleep, energy level and fatigue have also been reported to be confounders and were also included in our logistic regression analysis (Stuss & Levine, 1995; Goldberg et al., 1996).

g) Data analysis & sample size calculation

The database was created based on the information collected from the clinical, psychological, and neuropsychological examinations using the program EPI-INFO. After that, a system file was be created in the program SPSS Plus Version 4.0 for posterior data analyses. The distribution of the data according to demographic characteristics and
the variables previously mentioned for the different groups were described. In order to analyze the association between different factors, cross-tabulations were produced by the different variables in the different subgroups. The agreement between our two pain scales (100mm VAS and patient self-assessment) were calculated by the overall percent agreement and the weighted Kappa Index, which measures agreement beyond chance agreement (Hennekens & Buring, 1987). Most of the measurements were on categorical scales, which are easy to understand, to administer, to analyze, and to score. Considering that the variables were mostly categorical or recoded, the difference in proportions among the subgroups and its associated variables were tested by means of the Pearson’s Chi-square and the Fisher’s exact test (two-sided test, p<0.05) as well as by odds ratio and 95% confidence interval. Odds ratio, Maentel-Henszel odds ratio, and 95% confidence interval were included, because the p-value as the single measure of association can be influenced not only by the magnitude of the difference but also by the sample size (Hennekens and Buring, 1987). If the data are continuous (e.g., reaction-time tests, the Brown-Peterson Consonant Trigram Auditory Memory Task, and the California Verbal Learning Test), one-way ANOVA, Student’s t-test and Paired Student’s t-test, Mann-Whitney U-Wilcoxon Rank Sum Test, and Tukey-b Multiple Range test (two-sided test, p<0.05) were employed. Finally, logistic regression were applied to the data once a full exploration has been conducted using simpler techniques. This is recommended for assessing the role of confounding for assessing internal validity and for discovering errors in an early stage of the analysis and prevents omission of important data (Schlesselman, 1982; Norusis, 1991, 1992).
h) Sample size calculation

In order to test our first hypothesis (difference between mean scores in the
europsychological tests between responding TMD (rTMD) and non-responding TMD
(nrTMD) patients, the formula for the calculation of the sample size for two independent
means (Taylor, 1981) was the following:

\[ n/group = 2 \left[ \left( Z_\alpha + Z_\beta \right) \sigma / \Delta \right]^2 \]; in which,

- \( n \) = the estimated sample size in each group,
- \( Z_\alpha \) = value of the standard normal distribution corresponding to a significance
  level of alpha (e.g., 1.96 for a two sided test at the 0.05 level),
- \( Z_\beta \) = value of the standard normal distribution corresponding to the desired level
  of power (e.g., 0.84 for a power of 80%),
- \( \sigma^2 \) = sample variance assuming equal variability in the two groups, and
- \( \Delta \) = assumed expected difference.

Considering the novel characteristic of this study, previous data on the
neuropsychological differences between the rTMD and nrTMD groups are nonexistent.
Due to this fact, the sample size was calculated using the means and standard deviations
between Groups I and II from the results of the simple time-reaction test (SRT) of our
preliminary data. This test was chosen, because it showed the smallest significant
difference between the two groups among all reaction-time tests \( P < .05 \). The
calculation showed that a total of 17 individuals are needed in Group II (nrTMD) in order to detect a difference of 75 mSEC between the two groups. Considering that the nrTMD is a subdivision of the initial TMD group and that no more than 30% (Greene & Laskin, 1983) of the initial TMD group improves after pain management (nrTMD, Group II), it was estimated that 57 patients must be screened in the initial TMD group in order to obtain a sample of 17 individuals in the nrTMD. The number was increased to 60 in the TMD group and to 20 in the nrTMD to compensate for drop-outs (Dao et al, 1994). The number of patients in the IBS and non-pain group were also 20 to allow the comparisons between groups of similar numbers. Recalculations using our own data reduced the final number from 17 to 14, which did not change our initial sample size estimation; therefore, no changes were made.

i) Examiner

In order to assess the clinical signs and symptoms involved in TMD, one experienced clinician in the area of orofacial pain was sufficient. A single examiner was chosen, because the intra-examiner Kappa index had been shown to be similar or higher than the inter-examiner reliability (0.62)(Dworkin 1990b). Carlsson et al. (1980) found that the intra-observer variation during a five-week period was smaller than the differences between two different observers in the same occasion. However, this intra-observer variation increased one year after treatment. She suggested that in longitudinal
studies, repeated registrations should be performed by the same observer, since the intra-
observer consistency seems to be greater than that between the different observers at
different times. However, this agreement decreased substantially one year after treatment.
Considering that the clinical examination was performed only at baseline, there was no
need for calibration. Besides, even if calibrated, after six months the fluctuation of signs
and symptoms would have made the intra-observer variation increased and compromised
the reproducibility of the results.

Finally, the examiner was blinded to both responding and non-responding TMDs, because the treatment outcome evaluation will take place 6 months after baseline
assessment; and therefore, the patients would be tested prior to even knowing the treatment
response. In addition, neuropsychological tests (SRT, MCRTs, CVLTs, and CCC) are
computer driven so the examiner, upon administrating the tests, was biased.

j) Informed consent, confidentiality and budget justification

The examination and interviews cover aspects of personal and family disease
histories, including sensitive facts such as educational level attained, income bracket, and
age. Therefore, the patient has the right to privacy and protection from exploitation of
delicate material, and the study has the patient's consent and the approval of the Human
Ethics Committee of the University of Toronto Office of Research Services. Twenty-five
dollars were paid to each participant in order to increase recruitment and decrease drop-
outs. The participation rate of the eligible patients was slightly over fifty percent. The major reason for this relatively low participation was the inconvenience of time and parking costs involved.

k) Pilot testing

Pilot testing with a small sample size was carried out prior to the initiation of the study per se for the following reasons: (a) it might uncover unnoticed problems; (b) it might correct improper phrasing of questions; (c) it might identify missing questions; (d) it might show that the questionnaire is too long and tiresome; (e) and provided the preliminary data for the final sample size calculation (Schlesselman, 1982). Our pilot testing showed no restrictions to our initial protocol which remained unchanged. The pilot sample size (n = 10) was included in our final database.

l) Protocol for the TMD group

The diagram of this investigation is summarized in Figure 1. The detailed description is described below in the text.
Figure 1: Diagram of the research protocol

(A) Diagnostic criteria: RDC/TMD (Dworkin & LeResche, 1992; Dao et al., 1994; Goldberg et al., 1996)
(B) Diagnostic criteria: Drossman, Li & Toner et al., 1995
(1) Neuropsychological (reaction time tests, CVLTs, CCC) and psychosocial assessment (sleep - SAQ, depression - BDI, fatigue and energy level - 100 mm VAS, educational level, employment income, age)
(2) Clinical history and examination (palpation of TMJs and masticatory muscles, pain intensity at rest and on chewing pre- and post-treatment, length of treatment, pain duration, number of treatments treating clinician, comorbidity with IBS, maximum mouth opening, overbite, overjet, percussion sensitivity, caries, and pain exacerbation after examination)
i) Patient selection

Patient selection and diagnosis was made in the first appointment at the Mount Sinai Orofacial Pain Clinic and in the TMJ Clinic at the Faculty of Dentistry by one of the four experienced treating clinicians. Selected patients were then asked about their willingness to participate in the study. The overall description of the study objectives, risks to the patient, compensation, confidentiality issues, and clarifications were provided. Those who agreed to participate were scheduled for a test appointment no earlier than 3 days and no later than 2 weeks. A cheque in the amount of $25.00 dollars was mailed by the secretary of the clinic to the patient's given address after both the neuropsychological and psychosocial evaluation had been completed and the follow-up questionnaire had been returned.

ii) Wash-out period and scheduling of appointments

The first three days were used as a wash-out period, when the patient should have discontinued the use of medication which might have influenced our test results. In addition, the patient should be willing to be tested within to weeks to prevent that the effect of the treatment might have influenced our test results. The patient was contacted also by phone in the prior evening to confirm the test appointment on the following day.

iii) Neuropsychological and psychosocial assessment

The meeting place was the dental clinic, because the access is easier than the neuropsychological test site. From there the patient was taken into the test room. All
patients were schedule for the same test site (Mount Sinai Hospital Craniofacial Pain Research Unit) between 10 am and 4 pm in order to prevent different test results due to different locations or pain daily variation (Folkard, 1976). The room was located in the basement which was the most silent floor in the hospital.

In the room, the patient was usually asked if she needed to leave the room for any reason for one hour; because the tests have to be ran without interruptions. Initially, the patient was presented with the consent form and instructed to carefully read, and she was also encouraged to ask questions. If the patient was under 18 years old, one of the parents was instructed to read and sign on her behalf. The door of the test site was kept opened at all times. The first neuropsychological test that the patient had to undergo were the reaction time tests in the order of increasing difficulty (simple, multiple-choice reaction time test, multiple choice reaction time test with conflict, and multiple choice reaction time test with constraint)(Stuss et al., 1987, 1989a, 1989b; Goldberg et al., 1996). All tests were standardized in the following way: 25% percent probability of the target object coming every time, each object stayed on the screen for 2 seconds, and there was a 5 second delay between one appearance and another. All tests were computer driven which almost eliminated the possibility of examiner or patient response bias, but a initial explanation was given to the patient about what the patient had to do before the beginning of each test. The examiner also asked the patient if she had any questions regarding the test. The test was only initiated after the patient reported no concerns.

The first test, the simple reaction time test (SRT), took exactly 4 minutes and 10 seconds and had 50 trials. In each trial, the patient had to press the only button she had as
soon as possible with the dominant hand whenever she saw the target geometric shape. The target could be either a circle, a square, a triangle, or a cross and was chosen by the computer at the beginning of each test in a random manner. The computer then scored the time in each trial and gave an average mean reaction time for all trials plus standard deviation.

All the other reaction time tests (multiple choice reaction time tests) took 8 minutes and 20 seconds with 2 minutes in between to set up the computer to another test. In the multiple choice reaction time test (MCRT), the patient had to press two buttons instead of one with both the dominant and non-dominant hand. Four different shapes (circle, square, triangle and cross) came at the same interval with a 25% probability of appearance. The target was chosen by the computer prior to the beginning of the test, and the patient had to press the button in his dominant hand as fast as possible whenever she saw the target. If the object on the screen was a non-target, than she pressed the button in her non-dominant hand.

In the multiple choice reaction time test with conflict (MCRTCF), the patient also had to press the button in the dominant hand whenever the target object appeared on the screen. However, the level of difficulty increased, because the target differs from non-targets not only in shape but also in color and in the orientation of the lines inside. For example, I told the patient "your target will be a yellow square with vertical lines inside. Any other similar object, such as a green square with vertical lines, or a yellow circle with vertical lines or even a yellow square with horizontal or oblique lines must be considered non-targets."
For the multiple choice reaction time test with constraint (MCRTCT, the last computer driven one), the same instructions given for the previous test were given to the patient. The only difference was that the computer never repeated the same shape with a different color, which made the test very similar to the MCRT. This was done also for internal validity, because it allowed us to compare the reaction times and to see if the speed in response was similar to what was expected according to the level of difficulty of the test in increasing order: SRT, MCRT, MCRTCT and MCRTCF.

Immediately after, the patient underwent the California Verbal Learning Tests. The test was administered by the examiner in a standardized manner and the same sixteen-list of words was read once during five trials. For the first trials, the patient was read the following instructions: “Let’s suppose you were going shopping on Monday. I’m going to read a list of items for you to buy. Listen carefully, because when I’m through, I want you to repeat as many of the items as you can. It doesn’t matter what order they are in - just tell me as many as you can.” The patient was then asked if she had any questions or concerns. For trials 2 - 5, the patient was read the following: “I’m going to repeat Monday’s shopping list. Again, I want you to repeat as many items as you can, in any order, including items you may have already told me.” Again, the patient was questioned if she understood the instructions. The total number of correct responses of trials 1 - 5 was scored in the correct responses sub-section (CVLTCR), and the number of semantic clusters (words reported already group according to categories; i.e. spices, clothes, tools and fruits in the same trial) was scored in the sub-section (CVLTCL). The total number of perseverations, number of items said twice or more in the same trial, were score in the
sub-section (CVLT-C) and the number of intrusions, the total number of reported words non-existing in the list in each trial, were also registered in the CVLTI final score.

After the CVLT, the patient was presented with the F.E.P.s Self-Completing Questionnaire. A questionnaire created in our research unit which included questions about fatigue, energy level, facial/jaw pain at rest, facial/jaw pain on chewing, educational level, employment and income level. This questionnaire was given in between the first two neuropsychological tests (i.e., reaction time tests and CVLT) and the last one (CCC). This was done to give the patient some rest and to prevent early fatigue which might have influenced the test results (Goldberg et al., 1996). The questions on fatigue, energy level, pain at rest and pain on chewing were all assessed by 100 mm VAS and all asked about the average level of pain, fatigue and energy level within the last month to prevent diurnal variations (Folkard, 1976). The VAS had a explanatory heading to make sure that the patient marked the scale correctly: “Please mark each line with an X”. The remaining variables were assessed by categorical scales: a) educational level (seven categories), employment (2 categories) and income level (six categories). For the purpose of statistical analysis, the different categories were collapsed into two. The patient was instructed to feel free to report his concerns regarding, not only the questions but also the proper way to fill out the F.E.P.s Self-Completing Questionnaire.

After the F.E.P.s Self-Completing Questionnaire, the patient was given the University of Toronto Sleep Assessment Questionnaire (Cesta, Moldofsky & Sammut,
1996). She was instructed to fill out date and date of birth prior to the beginning of the test. Then the patient read the following instructions on the SAQ: “Please answer each question by question by checking (✓) the ONE answer that fits best.” The patient was then questioned if she understood how to fill out the SAQ. If yes, she answered the 19-item questionnaire according to the instruction.

Following the SAQ, the patient is presented with the Beck Depression Inventory (BDI) for assessment of depression. The 21-item questionnaire, similar to the SAQ, was also self-completing, but it had no questions, only four choices numbered 0 to 3. The patient was instructed to read carefully each of the alternatives and choose one. After that, she had to put the number corresponding to her choice inside the box for responses in each question. Similar to the other questionnaires, the patient was asked if she had any questions or concerns before filling out the BDI.

The last neuropsychological test was the Brown-Peterson Consonant Trigram Auditory Memory Task (CCC). The test was administered by the examiner in a standardized manner to minimize bias. The first five items were not scored and were used as a warm-up. Before initiating the warm-up test, the patient was read the following instructions: “I am going to say three letters and when I am through. I am going to knock like this (examiner demonstrates). When I do, I want you say the letters back.” Then, before initiating the final 15 actual questions, the patient was instructed with the following: “This time, I am going to say three letters followed immediately by a number. As soon as you get the number, I want you to start counting backward by three’s out loud
like this 89-86-83. Continue counting out loud until I knock as before. (Examiner demonstrates knocking on the desk). When I knock, I want you to recall the three letters. Do you have any questions?” Then, using a chronometer to detect the 3, 9 and 18-second delay between counting and remembering the trigram (the three letters), the test began.

iv) Clinical examination & treatment regimen

All groups underwent the neuropsychological tests, and only the TMD group was submitted to clinical examination. After that, the TMD patients underwent clinical examination. Clinical examination was performed last, because the pain exacerbation after examination could have influenced our test results. The patient was then conducted to a dental office already prepared for clinical examination. The clinical examination was then performed in the following order: a) palpation of the TMJ (lateral pole or facial palpation, posterior ligament inside the external auditory meatus, and sounds), b) palpation of the masticatory muscles extraorally (masseter, temporalis and sternocleidomastoid) and intraorally (medial pterygoid area, lateral pterygoid area, and insertion of the temporalis), c) maximum unassisted mandibular opening, d) overbite, e) overjet, f) percussion sensitivity, g) dental caries, and h) verbal inquiry regarding exacerbation of pain from examination.
Patients with TMD underwent reversible procedures: a) lower flat bite-plane including full coverage in the posterior teeth, partial coverage (cingulum line) in the anterior teeth, no canine guidance and were made of transparent hard acrylic resin, b) low dose muscle relaxants - cyclobenzaprine (Flexeril, 5 to 10 mg, one at bedtime), c) non-steroid anti-inflammatory drug - diflunisal (Dolobid, 500 mg, four times a day), and d) physical therapy (moist heat, massage and ultrasound therapy). The patients were reassessed after treatment using a mailed follow-up questionnaire for pain intensity at rest and pain on chewing post-treatment (100 mm VAS) as well as for subjective pain intensity assessment (better, same, worse). The patients who did improve after reversible treatment constituted Group I (rTMD), and those who did not were part of Group II (nrTMD).

The baseline measures between rTMD and nrTMD were compared and used to assess the prognostic utility of the neuropsychological tests, clinical examination, and psychological and behavioral characteristics.

v) Follow-up questionnaire

The follow-up questionnaire was mailed out six months after initial appointment at the Orofacial Pain Clinic at the Mount Sinai Hospital and TMJ Clinic at the Faculty of Dentistry. The follow-up questionnaire was self-completing and contained questions about pain intensity at rest within the last month, pain on chewing at rest within the last
month, and patient's self-assessment of pain improvement. The format of the first two questions were identical to the one assessed at baseline. Only the question regarding patient's self-assessment was exclusive of the follow-up questionnaire. The format of each question was described in the improvement criteria section.

m) Protocol for IBS patients

The protocol for IBS was very similar to the TMD group. The differences was that the IBS contacted the examiner by phone. The sources of IBS patients were major rheumatology outpatient clinics in the Toronto Western Hospital, Toronto General Hospital and Mount Sinai Hospital. Advertisements were left with the secretary of each clinic to be given to non-responding IBS patients selected by the treating specialist. The clinician followed our inclusion/exclusion criteria described in the IBS inclusion/exclusion criteria section. In the advertisement, general information about the study was given as well as the examiner's home phone number for contact. Once contacted, the examiner provided extra clarifications if needed and scheduled the appointment if the patient agreed to participate in the study.

In addition, a newspaper advertisement was posted recruiting IBS patients in a major newspaper which served Metro Toronto. Patients who responded to the add were given the same basic information given in writing to the IBS group referred from specialists. The only difference was that the patient must give the name and phone
number of treating clinician. The name and phone number was checked in the list of the Canadian Medical Association Directory. The physician was then contacted over the phone for confirmation.

Patient's time scheduling and neuropsychological testing were the same described for the TMD group, with the exception that IBS patients did not undergo clinical examination. Different than the IBS group, the payment was mailed out after the neuropsychological and psychosocial evaluation, considering that only the TMD group underwent follow-up assessment.

n) Protocol for non-pain patients

Non-pain patients were recruited from advertisements posted throughout the University of Toronto Main Campus. In addition, advertisements were posted in the Mount Sinai Hospital Orofacial Pain Clinic as well as in Faculty of Dentistry. The volunteers who contacted the examiner were given the same information given to the TMD and IBS groups. The patient was inquired about his medical history to evaluate if they matched our inclusion/exclusion criteria already described. If the patient agreed to be a volunteer, she was scheduled an appointment. Similar to the IBS group, only the neuropsychological and psychosocial testing were performed and the payment procedures were identical.
VI. RESULTS

In general, my findings suggested that a cohort of neuropsychological tests were highly predictive of treatment failure as outlined below. The results and analyses were presented in four section: a) descriptive data on recruitment rates, b) assessment of missing values, c) sociodemographic description of the populations studied at baseline, d) association between verbal and non-verbal improvement criteria, e) bivariate analysis - differences between nrTMD and rTMD, f) determination, using 2x2 tables and logistic regression, of the best predictors of TMD treatment outcome controlling for confounders (hypothesis 1), g) determination of the similarity or lack thereof of neuropsychological test scores obtained from nrTMD and IBS patients versus rTMD and IBS patients using 2x2 tables and logistic regression (hypothesis 2), h) distribution of the original values of the neuropsychological tests and confounders between and among the four groups, and i) backward step logistic regression analysis in order to select the best predictors of TMD treatment outcome controlling for many variables at the same time.

a) Descriptive data on recruitment and follow-up rates

The log of all new orofacial pain patients recruited between June 01 and August 13, 1996 (n = 116) is shown in Appendix I. Only 19% (n = 22) of the total number of new patients met the inclusion/exclusion criteria, and of these, only 50% agreed to participate
in the study. This decreased the available sample to 9 new patients every approximately 3 months or 3 new patients every month. Thus, two years were required for data collection to be completed (n=60). The primary diagnosis was myofascial pain with or without TMJ disk interference disorders. Besides TMD, other common diagnoses were post-traumatic TMD due to Motor Vehicle Accidents (MVA) (10%), neurological disorders (9%), headaches (8%), TMJ pain without muscle involvement (7%), atypical facial pain (7%), strictly psychological disorders (7%), oral mucosal diseases (6%), among others. Only one patient which did not meet our inclusion criteria regarding age (15 to 45 years old) was included. The patient was a thirteen year old TMD patient with a diagnosis of myofascial pain. Considering that this small age difference was not shown to be different than other age groups for neuropsychological tests (Stuss et al., 1987), the patient was considered to be eligible. By inspection, her scores did not differ from the TMD sample means for neuropsychological tests and psychosocial variables. The follow-up rate was 100%, which was considered excellent.

b) Assessment of missing values

In Appendix II, the number and proportion of missing values are reported. The vast majority of the missing values for the 45 variables tested did not reach 5%. The variables which have a proportion of missing values greater than 5% are the following: a) pain duration (17.6%), comorbidity with irritable bowel syndrome (11.1%), age (10%), number of treatments provided (9%), length of treatment (9%), and treating clinician
(9%). These variables were then compared among the two or four groups, considering that only age was assessed for the four groups. The different proportions of missing values across the two or four groups was compared using Chi Square test for difference between proportions. No statistically significant difference was found in the distribution of the missing values tested across the two or four groups (Appendix III).

c) Sociodemographic description of the populations studied at baseline

The sociodemographic variables are examined in Table I to assess if there were significant differences at baseline among the three groups (TMD, IBS and non-pain population). Only three groups were compared, because at baseline we were still blinded to the TMD outcome, and thus the nrTMD and rTMD groups could not be identified at this point. No significant difference was found in the variables assessed: a) educational level, b) employment, c) income, and d) age. No significant differences were found among or between the three groups for all variables (Chi Square test for differences in proportions and one-way ANOVA for differences in means, $p > 0.05$). The majority of the population studied for all groups had a high degree of education (post-secondary diploma/certificate or higher). The majority of the TMD population was employed (67.8%); for the IBS group, 50% were employed; while in the non-pain population only 40% were employed. For all three groups, the majority of the population belonged to low income group ($39,000 a year or less). The mean age for all groups was very similar including the age range, and it was approximately 30 years with a range of 12 - 46.
d) Verbal versus non-verbal improvement criteria

Both the verbal and non-verbal improvement criteria as well as percent agreement between them and Kappa index are shown in Table II. In the non-verbal improvement criteria (30% or greater reduction in the 100 mm VAS of pain at rest over baseline) versus the verbal improvement criteria (better, same or worse), the overall percent agreement was 90% and the Kappa index controlling for observer agreement by chance (set at 50%) was 0.78. This level of agreement is considered by Landis & Koch (1977) to
be substantial (0.61 - 0.80). Sixty percent of TMD patients showed a 30% reduction in the VAS in pain at rest over baseline (non-verbal criteria), while sixty three percent in the verbal criteria showed improvement.

### TABLE II: ASSOCIATIONS BETWEEN VERBAL AND NON-VERBAL PAIN IMPROVEMENT

<table>
<thead>
<tr>
<th>Independent Variables: Unit or category</th>
<th>Non-verbal improvement (%)</th>
<th>Non-verbal improvement (%)</th>
<th>Overall percent agreement</th>
<th>Kappa index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Thirty percent reduction or greater in VAS* in pain at rest over baseline = 0</td>
<td>Less than thirty percent reduction in VAS in pain at rest over baseline = 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal improvement: better = 0 (n = 38)</td>
<td>94.4 (n = 34)</td>
<td>16.7 (n = 4)</td>
<td>90.0</td>
<td>0.78</td>
</tr>
<tr>
<td>Verbal improvement: same or worse = 1 (n = 22)</td>
<td>5.6 (n = 2)</td>
<td>83.3 (n = 20)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Visual analogue scales (100 mm)

e) Determining the best TMD treatment outcome predictors (Hypothesis I)

Differences in means between Groups I (rTMD) and II (nrTMD) for the neuropsychological psychosocial tests, confounders, tmj and masticatory muscle palpation, occlusal variables and pain intensity are shown in Tables III - VIII.

Student’s t-test and the Mann-Whitney U-Wilcoxon Rank Sum Test were used for continuous variables. If the variables were categorical, categorical data analyses (2 x 2 tables) using Chi Square and Fisher’s Exact tests for determining statistically significant
difference (p<0.05) between the two groups; as well as odds ratio (critical OR=2.0) and 95% confidence interval to determine the strength of the associations were used. The role of confounders was calculated by variation (15%) in odds ratio using logistic regression and the effect of interactions was compared by variation in the beta coefficient (15%). The strength of association with predictors measured originally as continuous variables, once they have already been controlled for confounders was assessed. They were dichotomized based on the literature and the present data (minus two standard deviations). When a cutoff point is selected, both sensitivity and specificity will be provided. Positive test results will be indicated as one (1) and negative test results will be indicated as zero (0). The nrTMD group (Group II) will be indicated as one (1), or equivalent to disease positive; and the rTMD group (Group I) will be indicated as zero (0), or equivalent to disease negative.

i) Neuropsychological tests

On the basis of the reaction time tests (scores 0 to infinite) (simple reaction time test - SRT, multiple choice reaction time test - MCRT, multiple choice reaction time test with conflict - MCRTCF and the multiple choice reaction time test with constraint - MCRTCT), there was no statistically significant difference between the two groups (Student’s t-test and Mann-Whitney U-Wilcoxon Rank Sum Test, p > 0.05)(Table III). In all tests, nrTMD was slower than rTMD, but the actual difference observed between the two groups in all reaction time tests was very small (range 12 - 44 msec) which is
substantially less than the value used for the sample size calculation (75 msec). Differences less than 100 msec were considered to be non-important by Stuss et al. (1985, 1989a, 1989b).

Results obtained from tests that evaluated attention and short-term memory under interference (California Verbal Learning Test correct responses- CVLTCR, CVLT clusters - CVLTCL, CVLT perseverations - CVLTP, CVLT intrusions - CVLTI, and the Brown-Peterson Consonant Trigram Auditory Memory Task - CCC) showed statistically significant differences between the two groups (Student's t-test, Mann-Whitney U-Wilcoxon Rank Sum Test, p<0.05)(Table III). The CVLTCR (scores 0 - 80) showed notable differences between the two groups (p<0.01), with nrTMD remembering fewer correct words in a shopping list (mean = 53, SD = 10) than rTMD (mean = 60, SD = 8.5). Similarly, there were very profound differences (p < 0.001) in the scores on the CVLTCL (0 - 60) between the two groups, with nrTMD patients grouping fewer words semantically (e.g. different spices, clothes, tools and fruits) (mean = 18, SD = 7.4) than rTMD patients (mean = 27, SD = 11). The other sub-sections of the CVLT, the CVLTI (scores 0 - 10) (number of intrusion of words which did not belong to the list) and the CVLTP (scores 0 - 40) (perseverations or repetitions of words in the list) did not demonstrated differences between the groups. When the Brown-Peterson Consonant Trigram Auditory Memory Task (CCC) (scores 0 - 45) was used, clear and significant differences (p < 0.01) between nrTMD and rTMD patients was shown, with nrTMD patients identifying fewer correct trigrams (group of three letters) (mean = 30, SD = 6.3) than rTMD patients (mean = 34, SD = 6.3).
ii) Psychosocial tests

The psychosocial tests (100 mm VAS for fatigue and energy level, Beck Depression Inventory and the Sleep Assessment Questionnaire) were also able to discriminate between the test groups. Similarly, the SAQ (scores 0 - 68) also showed differences between the two groups, with nrTMD patients having more sleep disorders (Student’s t-test, Mann Whitney U-Wilcoxon Rank Sum Test, p < 0.05) (mean = 24, SD = 6.8) than patients with rTMD (mean = 20, SD = 6.2). According to the Beck Depression Inventory (scores 0 - 63), there were no statistically significant differences between the groups although there was a trend toward more depression in nrTMD (mean = 11, SD = 6.9) as compared to patients with rTMD (mean = 7.8, SD = 6.4). One interesting observation is the fact that fatigue and energy level, despite being closely related phenomena, were not shown to be perfect corollaries of one another. Fatigue (100 mm VAS) was higher for nrTMD (mean = 68, SD = 25) than rTMD patients (mean = 46, SD = 27), and with a highly statistically significant difference (p < 0.01) between the two. On the other hand, energy level (100 mm VAS), despite higher for patients with rTMD (mean = 50, SD = 24) than those with nrTMD (mean = 44, SD = 26), it was non-significant.
**TABLE III: NEUROPSYCHOLOGICAL AND PSYCHOSOCIAL TEST RESULTS BETWEEN RESPONDING (GROUP I) AND NON-RESPONDING TMD (GROUP II) PATIENTS**

<table>
<thead>
<tr>
<th>Independent Variables: Unit or category</th>
<th>Responding TMD, Group I (n=36) Mean (SD)</th>
<th>Non-responding TMD, Group II (n=24) Mean (SD)</th>
<th>Student's t-test</th>
<th>Mann-Whitney Rank Sum Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Reaction Time: scores (0-∞) msec</td>
<td>249 (60)</td>
<td>261 (67)</td>
<td>p = 0.5; NS</td>
<td>p = 0.3; NS</td>
</tr>
<tr>
<td>Multiple Choice Reaction Time: scores (0-∞) msec</td>
<td>437 (61)</td>
<td>477 (92)</td>
<td>p = 0.07; NS</td>
<td>p = 0.1; NS</td>
</tr>
<tr>
<td>Multiple Choice Reaction Time with Conflict: scores (0-∞) msec</td>
<td>484 (73)</td>
<td>528 (107)</td>
<td>p = 0.09; NS</td>
<td>p = 0.1; NS</td>
</tr>
<tr>
<td>Multiple Choice Reaction Time with Constraint: scores (0-∞) msec</td>
<td>447 (66)</td>
<td>480 (99)</td>
<td>p = 0.2; NS</td>
<td>p = 0.2; NS</td>
</tr>
<tr>
<td>CVLT-CR: scores 0-80</td>
<td>60 (8.5)</td>
<td>53 (10)</td>
<td>p = 0.005**</td>
<td>p = 0.005**</td>
</tr>
<tr>
<td>CVLT-CL: scores 0-60</td>
<td>27 (11)</td>
<td>18 (7.4)</td>
<td>p = 0.001***</td>
<td>p = 0.001***</td>
</tr>
<tr>
<td>CVLT-P: scores 0-40</td>
<td>5.3 (6.2)</td>
<td>5.6 (4.2)</td>
<td>p = 0.8; NS</td>
<td>p = 0.8; NS</td>
</tr>
<tr>
<td>CVLT-I: scores 0-10</td>
<td>0.6 (1.0)</td>
<td>0.7 (1.4)</td>
<td>p = 0.7; NS</td>
<td>p = 0.7; NS</td>
</tr>
<tr>
<td>CCC: scores 0-45</td>
<td>34 (6.3)</td>
<td>30 (6.3)</td>
<td>p = 0.006**</td>
<td>p = 0.006**</td>
</tr>
<tr>
<td>SAQ: scores 0-68</td>
<td>20 (6.2)</td>
<td>24 (6.8)</td>
<td>p = 0.02*</td>
<td>p = 0.02*</td>
</tr>
<tr>
<td>BDI: scores 0-63</td>
<td>7.8 (6.4)</td>
<td>11.0 (6.9)</td>
<td>p = 0.08; NS</td>
<td>p = 0.08; NS</td>
</tr>
<tr>
<td>FATIGUE: VAS 0-100 mm</td>
<td>46 (27)</td>
<td>68 (25)</td>
<td>p = 0.004**</td>
<td>p = 0.004**</td>
</tr>
<tr>
<td>ENERGY LEVEL: VAS 0-100 mm</td>
<td>50 (24)</td>
<td>44 (26)</td>
<td>p = 0.4; NS</td>
<td>p = 0.4; NS</td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01, *** p < 0.001
iii) Assessing the role of confounders

From Table IV, nine potential confounders which could not be eliminated during the design stage (i.e. educational level, employment, income, age, length of treatment, pain duration, number of treatments after initial visit, improvement according to the treating clinician and comorbidity) were analyzed. Categorical data were analyzed by Chi Square and Fisher’s Exact test and continuous variables by the Student’s t-test (p < 0.05). Some important variables which could not be analyzed were types of treatment provided, litigation. The types of treatment provided could not be analyzed, because the four different reversible treatments employed in this study were not given in the same number and sequence. Any analysis would have to divide too much our two groups, which would have resulted in loss of statistical power. However, other related variables (e.g. success rate among the treating clinicians as well as length and number of treatments provided) will be used instead. If in our bivariate and logistic regression analyses, these variables do not affect our treatment outcome, than it will be unlikely that types of treatment would. The issue of litigation was addressed, considering that TMD patients with history of MVA were excluded, which are usually the ones more involved in medical and legal disputes.

Educational level for the purpose of analysis was recoded into two groups: a) some education after high school or less, and b) post-secondary diploma/certificate or higher. This was done, so that we could calculate the odds ratio. Educational level, which showed no statistically significant difference at baseline, also showed no statistically significant difference between the two TMD groups at follow-up. Nevertheless, rTMD
patients had a higher proportion of individuals with post-secondary education or higher (74.3%) than nrTMD ones (50%). The odds ratio was higher (2.8) than our established cutoff of 2.0; therefore, despite not being statistically significant, it might have played a role in the neuropsychological differences described on Table III. The wide 95% confidence interval (0.9 - 8.6) might have been a reflection of our limited sample size which might also have influenced our p value (p = 0.06) which was almost significant.

Employment was dichotomous variable and there was no need for recoding. Employment, similar to educational level, showed no statistically significant difference both at baseline and at follow-up. Notwithstanding, rTMD had a higher proportion of employed individuals (74.3%) than nrTMD patients (58.3%). The Odds Ratio (2.0) did not exceeded the critical value and was not considered relevant. In addition, the confidence interval was also wide (0.6 - 6.2); therefore, it was not considered a confounder.

The household income level was recoded as following: a) $40,000 or more, negative test result, and b) $39,000 or less, positive test result. Contrasting the previous confounders, a very highly statistically significant difference (p<0.0001) was found for income level at follow-up, which differed from our non-significant finding at baseline. The odds ratio were extremely high (7.6) as well as the 95% confidence interval (2.1 - 27.6). The proportion of subjects with low income level for nrTMD (83.3%) was significantly higher than for rTMD (39.4%). Income, similar to educational level, may have played a role in the neuropsychological test results.
Another confounder which may have played a role in our test results was length of treatment. The length of treatment for the nrTMD group was significantly higher (mean = 21.9 weeks, SD = 20.3; Student’s t-test, p < 0.01) than for rTMD patients (mean = 11.6 weeks, SD = 6.4). Therefore, in rTMD patients the treatment results remained stable for three months. Despite not influencing the results positively, there was the possibility that length of treatment may have influenced our results in the opposite direction and was included in our final analysis. Analogously, the number of treatments provided was also statistically significant (p < 0.05) and was also considered to be a possible confounder. Number of treatments provided was greater for nrTMD (mean = 2.3, SD = 0.6) than for patients with rTMD (mean = 1.8, SD = 0.8) and was statistically significant (p < 0.05).

The final confounder which may have played a role was comorbidity with IBS patients. In 27.3% (six out of 22 valid cases) of the nrTMD patients, previous treatment for IBS was provided; while, this was the case in only 6.5% (two out of 32 valid cases) of the rTMD group. The difference was marginally significant (p = 0.05), and a higher sample size would have disclosed it. Finally, the OR was also high (5.4), which was substantially higher than our critical level.

The remaining variables (i.e. age, pain duration and treating clinician) were not considered to be confounders. Analogous to baseline, the average age for rTMD was slightly higher (mean = 29.4 years, SD = 9.0) than nrTMD patients (mean = 26.7 years, SD = 9.0) but non-significant at follow-up. Age, however, is usually considered to be a confounder in most epidemiological studies; therefore, it will also be included (Hennekens & Buring, 1987). Pain duration was highly variable in both groups, and had
and average of 47.4 months (SD = 53.8) for rTMD and 41.6 months (SD = 45.7) for patients with nrTMD. The difference was very small and non-significant. Four different treating clinicians using similar reversible therapies (muscle relaxant, non-steroid anti-inflammatory drugs, bite planes and physiotherapy) had similar success rates (range 50% - 61.9%) and no statistically significant difference was detected.
### TABLE IV: CONFOUNDERS, RESULTS BETWEEN RESPONDING (GROUP I) AND NON-RESPONDING TMD (GROUP II) PATIENTS

<table>
<thead>
<tr>
<th>Independent Variables:</th>
<th>Responding TMD, Group I = 0 (n=36)</th>
<th>Non-responding TMD, Group II = 1 (n=24)</th>
<th>Odds Ratio (OR)</th>
<th>Test of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit or category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Educational Level (%):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post-secondary diploma/ certificate or higher = 0</td>
<td>74.3</td>
<td>50.0</td>
<td>2.8</td>
<td>p = 0.06 §; NS</td>
</tr>
<tr>
<td>some education after high school or less = 1</td>
<td>25.7</td>
<td>50.0</td>
<td>(0.9 - 8.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Employment (%):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>employed = 0</td>
<td>74.3</td>
<td>58.3</td>
<td>2.0</td>
<td>p = 0.19 §; NS</td>
</tr>
<tr>
<td>unemployed = 1</td>
<td>25.7</td>
<td>41.7</td>
<td>(0.6 - 6.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Income (%):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$40,000 or more = 0</td>
<td>60.6</td>
<td>16.7</td>
<td>7.6</td>
<td>p = 0.000***§</td>
</tr>
<tr>
<td>$39,000 or less = 1</td>
<td>39.4</td>
<td>83.3</td>
<td>(2.1 - 27.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Age:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>29.4 (9.0)</td>
<td>26.7 (9.0)</td>
<td>p = 0.2 §§§; NS</td>
<td></td>
</tr>
<tr>
<td><strong>Length of treatment (weeks):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>11.6 (6.4)</td>
<td>21.9 (20.3)</td>
<td>p = 0.009***§§§</td>
<td></td>
</tr>
<tr>
<td><strong>Pain duration (months):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>47.4 (53.8)</td>
<td>41.6 (45.7)</td>
<td>p = 0.6 §§§; NS</td>
<td></td>
</tr>
<tr>
<td><strong>Number of treatments:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.8 (0.8)</td>
<td>2.3 (0.6)</td>
<td>p = 0.02* §§§</td>
<td></td>
</tr>
<tr>
<td><strong>Treating clinician (%) improvement:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinician 1</td>
<td>50.0</td>
<td>50.0</td>
<td>p = 0.9 §; NS</td>
<td></td>
</tr>
<tr>
<td>Clinician 2</td>
<td>61.9</td>
<td>38.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinician 3</td>
<td>60.0</td>
<td>40.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinician 4</td>
<td>50.0</td>
<td>50.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Comorbidity (Irritable Bowel Syndrome - IBS) (%):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never been treated = 0</td>
<td>93.5</td>
<td>72.7</td>
<td>5.4</td>
<td>p = 0.05 §§; NS</td>
</tr>
<tr>
<td>Have been treated = 1</td>
<td>6.5</td>
<td>27.3</td>
<td>(0.9 - 30.1)</td>
<td></td>
</tr>
</tbody>
</table>

*$ p<0.05$, **$ p<0.01$, ***$ p<0.001$

$\S$ Chi Square test for differences between proportions, $\S\S$ Fisher’s Exact test for differences between proportions, $\S\S\S$ Student’s t test for differences between means, Critical Odds Ratio = 2.0
iv) Clinical signs and symptoms: TMJ and masticatory muscle palpation

One of the most common factors analyzed during the examination of TMD patients are TMJ and masticatory muscle palpation (Zarb et al., 1994). From Table V, we found only one predictor of treatment outcome: palpation of the posterior ligament of the TMJ in the external auditory meatus (TMJEAM). Responding TMD patients had a smaller proportion (40%) of individuals with tenderness to palpation of the TMJ posterior ligament than the ones with nrTMD (73.9%). The odds ratio (OR) were high (4.2) and the difference was statistically significant (p < 0.05, Chi Square test). The other confounders which achieved odds ratio equal or higher than our critical odds ratio (2.0) were the medial pterygoid (5.5); lateral pole of the TMJ (TMJ Facial) (2.4); combined joint score (2.1); and combined muscle score (2.0); however, the findings were non-significant.

All the other variables on Table V did not reach our critical OR and were not statistically significant. However, some differences between responding versus non-responding TMD patients will be reported. In the palpation of the lateral pole of the temporomandibular joint, non-responding TMD patients had a higher proportion of positive responses (78.3%) than the responding group (60%). Almost no absolute difference was found for TMJ sounds between rTMD and nrTMD patients (34% versus 39%, respectively). Similarly, for the masseter both groups had comparable values (77% versus 73%, respectively). For the temporalis, nrTMD patients had slightly higher
proportion of positive responses (65%) when compared to those with rTMD (54%). On the other hand, for the sternocleidomastoid, a higher proportion of patients in the rTMD group (42%) was positive to palpation than those in the nrTMD (34%). For the lateral pterygoid, coronoid insertion of the temporalis and muscle score, all the results were very similar. For all areas, nrTMD patients score slightly higher (95 - 100%) than rTMD ones (85 - 91%).
<table>
<thead>
<tr>
<th>Independent Variables:</th>
<th>Responding TMD, Group I = 0 (n=31) (%)</th>
<th>Non-responding TMD, Group II = 1 (n=16) (%)</th>
<th>Odds Ratio (95% Confidence Interval)</th>
<th>Test of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TMJ Facial:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>40.0</td>
<td>21.7</td>
<td>2.4</td>
<td>p = 0.14; NS §</td>
</tr>
<tr>
<td>from II through III = 1</td>
<td>60.0</td>
<td>78.3</td>
<td>(0.7 - 7.9)</td>
<td></td>
</tr>
<tr>
<td><strong>TMJ EAM:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>60.0</td>
<td>26.1</td>
<td>4.2</td>
<td>p = 0.01 §*</td>
</tr>
<tr>
<td>from II through III = 1</td>
<td>40.0</td>
<td>73.9</td>
<td>(1.3 - 13.4)</td>
<td></td>
</tr>
<tr>
<td><strong>TMJ Sounds:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>65.7</td>
<td>60.9</td>
<td>1.2</td>
<td>p = 0.7; NS §</td>
</tr>
<tr>
<td>from II through III = 1</td>
<td>34.3</td>
<td>39.1</td>
<td>(0.4 - 3.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Masteter:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>22.9</td>
<td>26.1</td>
<td>0.8</td>
<td>p = 0.7; NS §</td>
</tr>
<tr>
<td>from II through III = 1</td>
<td>77.1</td>
<td>73.9</td>
<td>(0.2 - 2.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Temporalis:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>45.7</td>
<td>34.8</td>
<td>1.5</td>
<td>p = 0.4; NS §</td>
</tr>
<tr>
<td>from II through III = 1</td>
<td>54.3</td>
<td>65.2</td>
<td>(0.5 - 4.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Sternocleidomastoid:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>57.1</td>
<td>65.2</td>
<td>0.7</td>
<td>p = 0.5; NS §</td>
</tr>
<tr>
<td>from II through III = 1</td>
<td>42.9</td>
<td>34.8</td>
<td>(0.2 - 2.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Medial Pterygoid:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>20.0</td>
<td>4.3</td>
<td>5.5</td>
<td>p = 0.1; NS §§</td>
</tr>
<tr>
<td>from II through III = 1</td>
<td>80.0</td>
<td>95.7</td>
<td>(0.6 - 48.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Lateral Pterygoid:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>14.3</td>
<td>0.0</td>
<td>1.7</td>
<td>p = 0.1; NS §§</td>
</tr>
<tr>
<td>from II through III = 1</td>
<td>85.7</td>
<td>100.0</td>
<td>(1.3 - 2.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Insertion Temporalis:</strong></td>
<td>(intra-oral)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>11.4</td>
<td>0.0</td>
<td>1.7</td>
<td>p = 0.1; NS §§</td>
</tr>
<tr>
<td>from II through III = 1</td>
<td>88.6</td>
<td>100.0</td>
<td>(1.3 - 2.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Muscle Score:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>8.6</td>
<td>4.3</td>
<td>2.0</td>
<td>p = 1.0; NS §§</td>
</tr>
<tr>
<td>from 7 through III = 1</td>
<td>91.4</td>
<td>95.7</td>
<td>(0.2 - 21.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Joint Score:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 0 through III = 0</td>
<td>42.9</td>
<td>26.1</td>
<td>2.1</td>
<td>p = 0.1; NS §</td>
</tr>
<tr>
<td>from 4 through III = 1</td>
<td>57.1</td>
<td>73.9</td>
<td>(0.6 - 6.6)</td>
<td></td>
</tr>
</tbody>
</table>

§ Chi Square test and §§Fisher’s Exact test for differences between proportions; * p < 0.05
Critical Odds Ratio = 2.0
v) Intra-oral examination and pain intensity

In Table VI, categorical data were analyzed by the Fisher’s Exact test for difference in proportions and continuous variables by the Student’s t-test and Mann-Whitney U-Wilcoxon Rank Sum test ($p < 0.05$).

No statistically significant difference between the rTMD versus nrTMD groups was found in any of the occlusal variables analyzed: maximum mouth opening, overbite and overjet. The absolute difference between the two groups was also very small: a) maximum mouth opening (43 mm versus 45 mm, respectively), b) overbite (35% versus 29%), and c) overjet (30% versus 30%).

Pain at rest pre-treatment also showed no difference between the rTMD versus nrTMD patients on the 100 mm VAS (64 mm versus 63 mm, respectively); however, pain at rest post-treatment show a very highly statistically significant difference (21 mm versus 60 mm, $p<0.001$) between them, demonstrating that there was actual reduction in pain levels with the treatment. The average proportion of improvement over baseline in rTMD patients was very high (67.1%; $p < 0.001$, Paired Student’s t-test). Conversely, pain on chewing pre-treatment showed a highly statistically significant difference ($p < 0.01$) between rTMD versus nrTMD patients on the 100 mm VAS (39 mm versus 60 mm, respectively) and demonstrated to be a good predictor. Parallel to pain at rest post-treatment, pain on chewing post-treatment also showed a high statistical significance ($p <$
0.001); however, differently from pain at rest, the degree of improvement over baseline was less pronounced (35%) but still significant (p < 0.01, Paired Student's t-test).

The other variables included in the intra-oral examination were dichotomous: a) percussion sensitivity, b) caries, and c) exacerbation of pain after examination. All variables were non-significant, and the OR did not reach our critical level (2.0). Most patients in both rTMD and nrTMD reported exacerbation of pain after examination (88.6% versus 91.3%, respectively); nevertheless, percussion sensitivity was negative in both groups (88.6% versus 82.6%) as was caries (100% both groups). Test for significance between proportions as well as odds ratio could not be computed for caries.
<table>
<thead>
<tr>
<th>Independent Variables:</th>
<th>Responding TMD, Group I (n=36)</th>
<th>Non-responding TMD, Group II (n=24)</th>
<th>Student's t-test</th>
<th>Mann-Whitney U-Wilcoxon Rank Sum test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean (SD)</td>
<td>mean (SD)</td>
<td>p = 0.3; NS</td>
<td>p = 0.2; NS</td>
</tr>
<tr>
<td>Max. Mouth Open (mm):</td>
<td>43 (6)</td>
<td>45 (7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overbite (mm):</td>
<td>35 (21)</td>
<td>29 (15)</td>
<td>p = 0.2; NS</td>
<td>p = 0.1; NS</td>
</tr>
<tr>
<td>Overjet (mm):</td>
<td>30 (15)</td>
<td>30 (15)</td>
<td>p = 1.0; NS</td>
<td>p = 0.8; NS</td>
</tr>
<tr>
<td>Pain at rest (100mm VAS): (pre-treatment)</td>
<td>64 (26)</td>
<td>63 (20)</td>
<td>p = 0.9; NS</td>
<td>p = 0.3; NS</td>
</tr>
<tr>
<td>Pain at rest (100mm VAS): (post-treatment)</td>
<td>21 (18)</td>
<td>60 (23)</td>
<td>p = 0.0001***</td>
<td>p = 0.0001***</td>
</tr>
<tr>
<td>Pain on chewing (100mm VAS): (pre-treatment)</td>
<td>39 (27)</td>
<td>60 (27)</td>
<td>p = 0.005**</td>
<td>p = 0.004**</td>
</tr>
<tr>
<td>Pain on chewing (100mm VAS): (post-treatment)</td>
<td>25 (20)</td>
<td>63 (28)</td>
<td>p = 0.000***</td>
<td>p = 0.000***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Variables:</th>
<th>Responding TMD, Group I = 0 (n=36)</th>
<th>Non-responding TMD, Group II = 1 (n=24)</th>
<th>Odds Ratio (95% Confidence Interval)</th>
<th>Fisher's Exact test for differences in proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percussion Sensitivity:</td>
<td>88.6 (11.4)</td>
<td>82.6 (17.4)</td>
<td>1.6 (0.3 - 7.3)</td>
<td>p = 0.7; NS</td>
</tr>
<tr>
<td>Caries:</td>
<td>100.0 (0.0)</td>
<td>100.0 (0.0)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Exacerbation:</td>
<td>11.4 (88.6)</td>
<td>8.7 (91.3)</td>
<td>1.3 (0.2 - 8.0)</td>
<td>p = 1.0; NS</td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01, *** p < 0.001; NA (not available)
vi) Assessing TMD predictors controlling for confounders

According to Hennekens & Buring, 1987, a confounding factor can be considered as a mixing of the effect of the exposure under study on the disease with that of a third factor. This third factor must be associated with the exposure and, independent of that exposure, be a risk factor for the disease. In such circumstances the observed relationship between the exposure and disease can be attributable, totally or in part, to the effect of the confounder. In this investigation, the “disease” was temporomandibular disorders and the “exposures” were the cognitive tests: Simple Reaction Time (SRT), Multiple Choice Reaction Time (MCRT), MCRT with conflict (MCRTCF), MCRT with constraint (MCRTCT), California Verbal Learning Test correct responses (CVLTCR), CVLT clusters (CVLTCL), CVLT perseverations (CVLTP), CVLT intrusions (CVLTI), and Brown-Peterson Consonant Trigram Auditory Memory Task (CCC). In our analysis, even the neuropsychological tests which did not show statistically significant differences between the two TMD groups were included, considering that a confounder might act not only creating false positives, but also false negatives (Norusis, 1991, 1992).

From Tables IV, we have already determined that age, educational and income level, comorbidity with IBS as well as length of treatment and number of treatments provided were associated with TMD treatment outcome and were included in our analysis. Other psychosocial variables which have been reported to be associated with the neuropsychological tests employed (e.g. fatigue, energy level, sleep and depression) were also included in the model (Goldberg et al., 1996). Finally, any variable which involved
pain experience (e.g. pain at rest pre-treatment, pain duration, and pain on chewing pre-treatment) were also included. In the total, thirteen potential confounders were included in our logistic regression analysis.

The presence or absence of a confounding factor should never be assessed by tests of statistical significance, because they are too influenced by sample size (Hennekens & Buring, 1987). In our analysis, we assessed the presence of confounders by logistic regression. Table VII shows the associations between these possible confounders and the neuropsychological tests. The neuropsychological tests were entered as continuous variables. The letter ‘y’ represents the odds ratio (log odds) of that particular neuropsychological test. The confounders were analyzed one at a time to assess the influence of each variable by the initial ‘y’ (odds ratio). The variation in the ‘y’ for each confounder was also recorded. We considered to be a confounder any variable which changes the initial ‘y’ value by more than 15% (Greenland, 1989). Once all variables were analyzed one by one, all of them were included at the same time in the model to assess changes in ‘y’. The odds ratio for all neuropsychological tests did not change significantly in any direction both for each individual variable and for all variables in combination. The range of percent correct observations was high (76.1 to 85.7%).

From Table VI, we determined that none of the thirteen confounders included in our logistic regression analysis influenced the association between neuropsychological test scores and treatment outcome. Therefore, the neuropsychological tests results described in Table III were not confounded by age, fatigue, energy level, sleep, depression, educational level, income, comorbidity with IBS, number of treatments
provided, length of treatment provided, pain duration, pain at rest pre-treatment, and pain on chewing pre-treatment or by all variables in combination. Accordingly, the CVLTCR, CVLTCL and the CCC were still good predictors for TMD treatment outcome after controlling for confounders.
<table>
<thead>
<tr>
<th>Confounder Test</th>
<th>y</th>
<th>Age</th>
<th>FATI</th>
<th>ENLV</th>
<th>SAQ</th>
<th>BDI</th>
<th>EDLV</th>
<th>INCO</th>
<th>NTX</th>
<th>LTX</th>
<th>IBS</th>
<th>PDUR</th>
<th>PREST PRTX</th>
<th>PCHEW PRTX</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRT</td>
<td>1.003</td>
<td>1.003</td>
<td>1.001</td>
<td>1.002</td>
<td>1.003</td>
<td>1.002</td>
<td>1.001</td>
<td>1.002</td>
<td>1.005</td>
<td>1.003</td>
<td>1.004</td>
<td>1.003</td>
<td>1.001</td>
<td>1.004</td>
<td></td>
</tr>
<tr>
<td>MCRT</td>
<td>1.007</td>
<td>1.007</td>
<td>1.006</td>
<td>1.007</td>
<td>1.005</td>
<td>1.007</td>
<td>1.007</td>
<td>1.006</td>
<td>1.007</td>
<td>1.010</td>
<td>1.008</td>
<td>1.008</td>
<td>1.006</td>
<td>1.006</td>
<td></td>
</tr>
<tr>
<td>MCRTCF</td>
<td>1.006</td>
<td>1.005</td>
<td>1.004</td>
<td>1.005</td>
<td>1.003</td>
<td>1.006</td>
<td>1.006</td>
<td>1.004</td>
<td>1.005</td>
<td>1.009</td>
<td>1.006</td>
<td>1.007</td>
<td>1.006</td>
<td>1.005</td>
<td></td>
</tr>
<tr>
<td>MCRTCT</td>
<td>1.005</td>
<td>1.005</td>
<td>1.004</td>
<td>1.004</td>
<td>1.002</td>
<td>1.005</td>
<td>1.005</td>
<td>1.005</td>
<td>1.004</td>
<td>1.009</td>
<td>1.005</td>
<td>1.007</td>
<td>1.005</td>
<td>1.007</td>
<td></td>
</tr>
<tr>
<td>CVLTCR</td>
<td>0.915</td>
<td>0.933</td>
<td>0.928</td>
<td>0.917</td>
<td>0.929</td>
<td>0.896</td>
<td>0.921</td>
<td>0.927</td>
<td>0.912</td>
<td>0.902</td>
<td>0.907</td>
<td>0.885</td>
<td>0.904</td>
<td>0.933</td>
<td>0.910</td>
</tr>
<tr>
<td>CVLTCL</td>
<td>0.901</td>
<td>0.891</td>
<td>0.903</td>
<td>0.898</td>
<td>0.907</td>
<td>0.888</td>
<td>0.902</td>
<td>0.919</td>
<td>0.899</td>
<td>0.888</td>
<td>0.892</td>
<td>0.897</td>
<td>0.893</td>
<td>0.911</td>
<td>0.783</td>
</tr>
<tr>
<td>CVLTP</td>
<td>1.014</td>
<td>1.002</td>
<td>1.024</td>
<td>1.021</td>
<td>1.056</td>
<td>1.043</td>
<td>1.010</td>
<td>1.055</td>
<td>1.015</td>
<td>1.008</td>
<td>0.992</td>
<td>0.996</td>
<td>1.013</td>
<td>1.012</td>
<td>1.055</td>
</tr>
<tr>
<td>CVLTI</td>
<td>1.106</td>
<td>0.942</td>
<td>1.174</td>
<td>1.150</td>
<td>1.211</td>
<td>1.168</td>
<td>1.093</td>
<td>1.218</td>
<td>1.002</td>
<td>1.157</td>
<td>1.003</td>
<td>1.234</td>
<td>1.110</td>
<td>1.266</td>
<td>1.004</td>
</tr>
<tr>
<td>CCC</td>
<td>0.886</td>
<td>0.888</td>
<td>0.862</td>
<td>0.867</td>
<td>0.876</td>
<td>0.884</td>
<td>0.869</td>
<td>0.892</td>
<td>0.890</td>
<td>0.829</td>
<td>0.893</td>
<td>0.906</td>
<td>0.883</td>
<td>0.882</td>
<td>0.761</td>
</tr>
</tbody>
</table>

* variation of odds ratio greater than 15% when compared to y. Variable codes: FATI (Fatigue), ENLV (Energy Level), SAQ (Sleep Assessment Questionnaire), BDI (Beck Depression Inventory), EDLV (Educational Level), INCO (Income), NTX (Number of Treatments), LTX (Length of Treatments), IBS (Comorbidity with IBS), PDUR (Pain Duration), PREST PRTX (Pain at Rest Pre-treatment), PCHEW PRTX (Pain on Chewing Pre-treatment)
vii) Odds ratio between TMD treatment outcome predictors controlling for confounders

One of the advantages of the calculations of the odds ratio is the fact that it gives the probability of the individual to develop a specific disease if he or she has the exposure of interest. In addition, it is also less influenced by the size of the sample than the p-value (Hennekens & Buring, 1987).

On Table VIII, we recoded the neuropsychological and psychosocial tests which have shown to be good predictors of TMD treatment outcome on Table III. Recoding procedures (cutoff points) was based either on the normative data of the tests or, when that is impossible due to lack of published norms, we used plus or minus two standard deviations from our non-pain population mean (Stuss et al., 1987).

The following tests were recoded: a) the California Verbal Learning Test correct responses (CVLTCR), the Brown-Peterson Consonant Trigram Auditory Memory Task (CCC), the University of Toronto Sleep Assessment Questionnaire (SAQ) and the Beck Depression Inventor (BDI). Depression was included despite not having reached statistical significance; nevertheless, the test had marginal significance (p = 0.08). In addition, more than one study has reported depression, or depression correlates, as a good predictor of TMD outcome (Gale & Funch, 1984; Gerschman et al., 1987; Fricton & Olsen, 1996). Ultimately, norms for the BDI have been published for both psychiatric and non-psychiatric population (Beck, 1970). Other tests which were positive predictors, such as the California Verbal Learning Test clusters (CVLTCL) and fatigue, were
excluded due to the fact that both do not have published norms and their high standard
deviations made it impossible to define a reasonable cutoff point.

In the CVLTCR, our internal control values (mean = 62, SD = 8.9) in the non-pain
group was exactly the same as the normative data (age and sex matched) published by
Delis et al. (1987)(mean = 62 - 64). However, the standard deviation was not reported
and our internal controls standard deviation were used. Two standard deviations of the
mean yielded a cutoff point equals 44. For the CCC, not only our internal control mean
values were almost the same as the external age and sex matched (Stuss et al., 1987), but
also our standard deviations (mean = 38, SD = 3.8 versus mean = 35, S.D. = 2.6,
respectively). This was in agreement with Stuss and coworkers findings that the CCC is
not significantly affect by variations in both age and gender.

For the SAQ, the normative data published by Cesta, Moldofsky & Sammut
(1996) were used. The cutoff point 16 was used, because it is the point with the highest
sensitivity (0.73) and one the highest specificities (0.80). A sensitivity and specificity of
0.75 is considered to be good by Dworkin & LeResche (1992). For the BDI, 15 was
chosen as the cutoff point, because it includes patients with no or minimal depression
according to Beck (1970).

From Table VIII, it can be determined that patients who have scores lower or
equal to 44 in the CVLTCR will have 3.4 times the probability of becoming non-
responding patients. The findings however were non-significant, because our sample size
was based on detecting differences in means rather than proportions, which would require
many more subjects (Norusis, 1991, 1992). For the CCC, the probability of becoming a
non-responding TMD patients increases 4.5 times if the patient has a score lower or equal to 30 (positive test result, minus two standard deviations). The difference between the two groups was highly statistically significant (p<0.01) despite the small sample size, but the 95% confidence interval was also wide (1.5 - 14.1). In the SAQ, patients who have a score greater or equal to 17 (positive test result, minus two standard deviations) will have 5 times more chances to develop a chronic pain condition and the findings were statistically significant (p<0.05), despite the wide confidence interval (1.1 - 25.3). In the BDI, similarly to Table III, were non-significant, but the probability of becoming a non-responding TMD patient increases 3.1 times if the patient has a score greater or equal to 16.

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were also calculated for all tests. The calculations were based on two disease populations (rTMD versus nrTMD). Scores of 0.75 can be considered good for all measures (Dworkin & LeResche, 1992). The CVLT CR had high specificity (0.94) but low sensitivity (0.16) and moderate PPV and NPV (0.66 and 0.63, respectively). The CCC had better results than the CVLT CR, with moderate to good sensitivity and specificity (0.58 and 0.76, respectively) as well as PPV and NPV (0.64 and 0.72, respectively). The results for the SAQ were excellent for the sensitivity (0.91) and NPV (0.85); however, they were considered low for the specificity (0.46) and PPV (0.47). Finally, the BDI had high specificity (0.88), moderate PPV and NPV (0.63 and 0.53, respectively), and low sensitivity (0.29). The implications of these findings will be discussed later.
TABLE VIII: ASSOCIATIONS BETWEEN TREATMENT OUTCOME PREDICTORS FOR TMD: NEUROPSYCHOLOGICAL (COGNITIVE) TESTS

<table>
<thead>
<tr>
<th>Independent Variables: Unit or category</th>
<th>Responding TMD, Group I = 0 (n=36) (%)</th>
<th>Non-responding TMD, Group II = 1 (n=24) (%)</th>
<th>Odds Ratio (OR) (95% Confidence Interval, CI)</th>
<th>Chi Square test for differences in proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVLT-CR: scores from 45 through 80 = 0</td>
<td>94.4</td>
<td>83.3</td>
<td>3.4</td>
<td>p = 0.15; NS</td>
</tr>
<tr>
<td>scores from 0 through 44 = 1</td>
<td>5.6</td>
<td>16.7</td>
<td>(0.6 - 20.2)</td>
<td></td>
</tr>
<tr>
<td>sens = 0.16, spec = 0.94</td>
<td></td>
<td></td>
<td>p = 0.66, NPV = 0.63</td>
<td></td>
</tr>
<tr>
<td>CCC: scores from 31 through 45 = 0</td>
<td>76.5</td>
<td>41.7</td>
<td>4.5</td>
<td>p = 0.007**</td>
</tr>
<tr>
<td>scores from 0 through 30 = 1</td>
<td>23.5</td>
<td>58.3</td>
<td>(1.5 - 14.1)</td>
<td></td>
</tr>
<tr>
<td>sens = 0.58, spec = 0.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPV = 0.64, NPV = 0.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAO: scores from 0 through 16 = 0</td>
<td>31.4</td>
<td>8.3</td>
<td>5.0</td>
<td>p = 0.03*</td>
</tr>
<tr>
<td>scores from 17 through 68 = 1</td>
<td>68.6</td>
<td>91.7</td>
<td>(1.0 - 25.3)</td>
<td></td>
</tr>
<tr>
<td>sens = 0.91, spec = 0.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPV = 0.47, NPV = 0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDI: scores from 16 through 63 = 0</td>
<td>88.2</td>
<td>70.8</td>
<td>3.1</td>
<td>p = 0.09; NS</td>
</tr>
<tr>
<td>scores from 0 through 15 = 1</td>
<td>11.8</td>
<td>29.2</td>
<td>(0.8 - 12.1)</td>
<td></td>
</tr>
<tr>
<td>sens = 0.29, spec = 0.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPV = 0.63, NPV = 0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01, ***p < 0.001
Critical Odds Ratio = 2.0
viii) Selecting TMD treatment outcome predictors using logistic regression

The multivariate technique for estimating the probability that an event occurs chosen in our study is the logistic regression model. Other techniques, such as multiple regression and discriminant analyses, pose difficulties when the dependent variable (TMD treatment outcome) can have only two values (improved or did not improve), which is the case in our study. Logistic regression requires fewer assumptions than discriminant analysis; and even when the assumptions required for discriminant analysis are satisfied, logistic regression still performs well (Hosmer and Lemeshow, 1989).

In logistic regression, like in other multivariate statistical analyses, it is possible to identify subsets of independent variables (neuropsychological and psychosocial) that are good predictors of the dependent variable (TMD treatment outcome). The logistic regression procedure has several methods available for model selection. The variables can be entered into the model one by one, or it is possible to use a forward stepwise selection or backward stepwise elimination for automated model building. In a forward stepwise selection, we initiate with a model that contains only the constant; and at each step, the variable with the smallest significance level for the score statistic, provided it is less than the chosen cutoff value (by default 0.05), is entered into the model. The process continues until no more variables are eligible to enter into the model. In the backward stepwise selection, the model starts with all variables and then, at each step, the variable with the largest significance level for the score statistic is removed, provided it is more than the chosen cutoff value (by default 0.05). The process continues until no more variable are
eligible to be removed from the model. In simple terms, is the opposite procedure of the forward selection (Norusis, 1991, 1992).

There is no perfect model selection method, and more than one model should be built. The selection was based on the model which had the best overall agreement and the highest sensitivity and specificity in the classification table (Norusis, 1992). In this investigation, the backward step elimination was chosen, because the model takes all the variables into account in the model and then eliminates the ones which are not significant, rather than working with a sub-set of variables adding the ones which are significant (Hosmer & Lemeshow, 1989).

The model used in this study included the neuropsychological variables which proved to be the best predictors for TMD (CVLTCR, CVLTCL and CCC)(Figure 2) as well as all the psychosocial (fatigue, energy level, SAQ and BDI) and potential confounders (e.g. educational level, income, length of treatment, pain duration, pain at rest pre-treatment, and pain on chewing pre-treatment) used on Table V. In multivariate techniques, such as in logistic regression, all these interdependent variables can be analyzed at the same time.

In the backward stepwise procedure, the best predictors were CCC, fatigue and length of treatment. The overall agreement was high (82.0 %) as well as the sensitivity (0.78) and specificity (0.85). The CCC proved to be a good unconfounded predictor in both bivariate and multivariate analyses. Fatigue and length of treatment were also good predictors in the bivariate analyses, but they have not been assessed for confounders
considering that this was not the objective of our study. Therefore, their use as predictors must be re-assessed in a study specifically designed to answer these questions.

FIGURE 2: RESULTS OF LOGISTIC REGRESSION ANALYSIS: BACKWARD STEP PREDICTORS OF TMD TREATMENT OUTCOME (NEUROPSYCHOLOGICAL-COGNITIVE TESTS, BEHAVIORAL VARIABLES, SLEEP AND CONFOUNDERS)

Dependent variable: TMD (n=50) treatment success (Yes = 1, No = 0)

Independent variables (n=10):

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig</th>
<th>R</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCC</td>
<td>-0.2666</td>
<td>.0883</td>
<td>9.1182</td>
<td>1</td>
<td>.0025</td>
<td>-3.221</td>
<td>.07660</td>
</tr>
<tr>
<td>FATIGUE</td>
<td>.3505</td>
<td>.1629</td>
<td>4.6313</td>
<td>1</td>
<td>.0314</td>
<td>.1959</td>
<td>1.4198</td>
</tr>
<tr>
<td>LENGTX</td>
<td>.1657</td>
<td>.0709</td>
<td>5.4687</td>
<td>1</td>
<td>.0194</td>
<td>.2249</td>
<td>1.1803</td>
</tr>
<tr>
<td>Constant</td>
<td>3.9579</td>
<td>2.3917</td>
<td>2.7384</td>
<td>1</td>
<td>.0980</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Percent correctly classified 82.00%
Sensitivity 0.78
Specificity 0.85
PPV 0.81
NPV 0.82
f) Associations between the rTMD and nrTMD versus the IBS groups

We will now proceed with our hypothesis two, trying to determine if the scores on the neuropsychological tests of the nrTMD are more similar to those of the group with irritable bowel syndrome than with the ones of the rTMD. We will start by comparing the scores between nrTMD patients versus those with irritable bowel syndrome. After that, we will compare the neuropsychological test results between rTMD versus IBS patients.

i) Neuropsychological and psychosocial tests: nrTMD versus IBS patients

In Table IX, the reaction time tests (scores 0 to infinite) (simple reaction time test - SRT, multiple choice reaction time test - MCRT, multiple choice reaction time test with conflict - MCRTCF and the multiple choice reaction time test with constraint - MCRTCT) showed no statistically significant difference between nrTMD and IBS patients (Student’s t-test and Mann-Whitney U-Wilcoxon Rank Sum Test, p > 0.05)(Table IX). In all tests, IBS patients had lower reaction time test results than those in the nrTMD group, but the actual difference observed between the two groups in all reaction time tests was very small and were not considered to be relevant (range 11 - 36 msec)(Stuss et al., 1985, 1989a, 1989b).

Some of the tests which evaluates attention and short-term and short-term memory under interference (CVLTCR, CVLTCL, CVLTP, CVLTI, and the CCC)
showed also no statistically significant difference between the two groups and their absolute difference in all tests was extremely small. Actually, in the CVLTCR (scores 0 - 80) the absolute value was exactly the same. In both the CVLTCL (scores 0 - 60) and the CCC (scores 0 - 45), the absolute difference was only one. In the CVLTI (scores 0 - 10) and the CVLTP (scores 0 - 40) the actual values were very low which allows greater variability, but still the actual difference in the two tests was approximately one.

The psychosocial tests (100 mm VAS for fatigue and energy level, Beck Depression Inventory and the Sleep Assessment Questionnaire) showed also similar results. In the SAQ (scores 0 - 68), nrTMD patients showed less sleep disorders (mean = 24, SD = 6.8) than the ones with IBS (mean = 26, SD = 8.3), but the actual difference was very small. In the BDI (scores 0 - 63), the levels of depression in IBS patients (mean = 15, SD = 9.7) was higher than nrTMD ones (mean = 11, SD = 6.9), but the difference was small and non-significant. Level of fatigue (100 mm VAS) were almost perfectly the same between the nrTMD (mean = 68, SD = 25) and IBS groups (mean = 65, SD = 24) and with no statistically significant difference between the two. On the other hand, energy level (100 mm VAS), despite higher for nrTMD patients (mean = 44, SD = 26) than those with IBS (mean = 29, SD = 19), was non-significant (Table IX).
**TABLE IX: NEUROPSYCHOLOGICAL TEST RESULTS BETWEEN NON-RESPONDING (GROUP II) AND IRRITABLE BOWEL SYNDROME (GROUP III) PATIENTS**

<table>
<thead>
<tr>
<th>Independent Variables:</th>
<th>Non-responding TMD, Group II (n=24)</th>
<th>IBS, Group III (n=20)</th>
<th>Student's t-test</th>
<th>Mann-Whitney U-Wilcoxon Rank Sum Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td>Simple Reaction Time:</td>
<td>261 (67)</td>
<td>241 (34)</td>
<td>p = 0.2; NS</td>
<td>p = 0.4; NS</td>
</tr>
<tr>
<td>scores (0--∞) msec</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Choice Reaction Time:</td>
<td>477 (92)</td>
<td>452 (47)</td>
<td>p = 0.2; NS</td>
<td>p = 0.5; NS</td>
</tr>
<tr>
<td>scores (0--∞) msec</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Choice Reaction Time with Conflict:</td>
<td>528 (107)</td>
<td>492 (56)</td>
<td>p = 0.17; NS</td>
<td>p = 0.3; NS</td>
</tr>
<tr>
<td>scores (0--∞) msec</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Choice Reaction Time with Constraint:</td>
<td>480 (99)</td>
<td>469 (71)</td>
<td>p = 0.7; NS</td>
<td>p = 0.9; NS</td>
</tr>
<tr>
<td>scores (0--∞) msec</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVLT-CR:</td>
<td>53 (10)</td>
<td>53 (9)</td>
<td>p = 0.9; NS</td>
<td>p = 0.9; NS</td>
</tr>
<tr>
<td>scores 0-80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVLT-CL:</td>
<td>18 (7.4)</td>
<td>19 (10)</td>
<td>p = 0.5; NS</td>
<td>p = 0.5; NS</td>
</tr>
<tr>
<td>scores 0-60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVLT-P:</td>
<td>5.6 (4.2)</td>
<td>4.3 (2.9)</td>
<td>p = 0.2; NS</td>
<td>p = 0.3; NS</td>
</tr>
<tr>
<td>scores 0-40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVLT-I:</td>
<td>0.7 (1.4)</td>
<td>1.4 (1.8)</td>
<td>p = 0.17; NS</td>
<td>p = 0.3; NS</td>
</tr>
<tr>
<td>scores 0-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCC:</td>
<td>30 (6.3)</td>
<td>31 (4.8)</td>
<td>p = 0.4; NS</td>
<td>p = 0.3; NS</td>
</tr>
<tr>
<td>scores 0-45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAQ:</td>
<td>24 (6.8)</td>
<td>26 (8.3)</td>
<td>p = 0.4; NS</td>
<td>p = 0.5; NS</td>
</tr>
<tr>
<td>scores 0-68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDI:</td>
<td>11 (69)</td>
<td>15 (97)</td>
<td>p = 0.07; NS</td>
<td>p = 0.11; NS</td>
</tr>
<tr>
<td>scores 0-63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FATIGUE:</td>
<td>68 (25)</td>
<td>65 (24)</td>
<td>p = 0.7; NS</td>
<td>p = 0.7; NS</td>
</tr>
<tr>
<td>VAS 0-100 mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENERGY LEVEL:</td>
<td>44 (26)</td>
<td>29 (19)</td>
<td>p = 0.06; NS</td>
<td>p = 0.08; NS</td>
</tr>
<tr>
<td>VAS 0-100 mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01, *** p < 0.001
ii) Confounders: nrTMD versus IBS

From Table X, four confounders which could not be controlled during the design stage (i.e. educational level, employment, income, age) were also analyzed. Categorical data were analyzed by Chi Square and Fisher’s Exact test and continuous variables by the Student’s t-test (p < 0.05). Clinical variables (e.g. pain intensity and pain duration) were analyzed, because IBS patients did not undergo clinical examination.

Educational level showed no statistically significant difference between the nrTMD and IBS groups. Notwithstanding, the nrTMD had a lower proportion of individuals with post-secondary education or higher (50%) than the IBS patients (75%).

Employment, similar to educational level, showed no statistically significant difference between the nrTMD and IBS groups. The difference between the two groups was only 8%. The household income level was recoded similarly to reported between the two TMD groups. Comparable to previous confounders, no statistically significant difference was found for income level. The proportion of subjects with low income level for nrTMD patients (83.3%) was higher than for those with IBS (65%) though. The average age for nrTMD (mean = 26.7, SD = 9.0) was lower than for IBS patients (mean = 32.9, SD = 10.5) and was marginally significant (p = 0.06). Apparently, none of the confounders have played a role in our neuropsychological and psychosocial test results; however, this finding still must be confirmed in our logistic regression analysis below.
<table>
<thead>
<tr>
<th>Independent Variables:</th>
<th>Non-responding TMD, Group II = 0 (n=24)</th>
<th>IBS, Group III = 1 (n=20)</th>
<th>Test of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational Level (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post-secondary diploma/</td>
<td>50.0</td>
<td>75.0</td>
<td>$p = 0.08$; NS</td>
</tr>
<tr>
<td>certificate or higher = 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>some education after high school or less = 1</td>
<td>50.0</td>
<td>25.0</td>
<td></td>
</tr>
<tr>
<td>Employment (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>employed = 0</td>
<td>58.3</td>
<td>50.0</td>
<td>$p = 0.5$; NS</td>
</tr>
<tr>
<td>unemployed = 1</td>
<td>41.7</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>Income (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$40,000 or more = 0</td>
<td>16.7</td>
<td>35.0</td>
<td>$p = 0.16$; NS</td>
</tr>
<tr>
<td>$39,000 or less = 1</td>
<td>83.3</td>
<td>65.0</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>26.7 (9.0)</td>
<td>32.9 (10.5)</td>
<td>$p = 0.06$; NS</td>
</tr>
</tbody>
</table>

* $p<0.05$, **$p<0.01$, ***$p<0.001$  
§ Chi Square test for differences between proportions  
$$§§$ Student's t test for differences between means

From Table X, we have already determined none of the confounders were associated with our findings from Table IX. On the total, seven confounders were included in our logistic regression analysis.

In our analysis, we assessed the presence of confounders by logistic regression for the same reasons described for the TMD groups. Table XI shows the associations between these possible confounders and the neuropsychological tests. The letter 'y' represents the odds ratio (log odds) of that particular neuropsychological test. Similar to Table VII, we considered to be a confounder any variable which changes the initial 'y'
value in more than 15% (Greenland, 1989). Similar to Table VII, the odds ratio for all neuropsychological tests did not change significantly with the exception of the CVLTI. In the CVLTI, the combination of all variables changed the test result. The rationale for this finding is also explained on Table VII. Analogously, the range of percent correct observations was also very high (71.7 to 79.4%).

From Table XI, we can conclude that none of the seven confounders included in our logistic regression analysis influenced the association between neuropsychological tests and treatment outcomes. Therefore, the neuropsychological tests results from Table IX were unconfounded for age, fatigue, energy level, sleep, depression, educational level and income. From our findings, we can assess that the neuropsychological test results between nrTMD and IBS groups were indistinguishable from each other.
TABLE XI: Odds Ratio and variation of odds ratio* of neuropsychological tests versus confounders: nrTMD (Group II) versus IBS (Group III)

<table>
<thead>
<tr>
<th>Confounder</th>
<th>y</th>
<th>Age</th>
<th>Fatigue Level</th>
<th>Energy Level</th>
<th>SAQ</th>
<th>BDI</th>
<th>Educational level</th>
<th>Income</th>
<th>All variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRT</td>
<td>0.992</td>
<td>0.991</td>
<td>0.993</td>
<td>0.992</td>
<td>0.992</td>
<td>1.008</td>
<td>0.992</td>
<td>0.991</td>
<td>0.987</td>
</tr>
<tr>
<td>MCRT</td>
<td>0.995</td>
<td>0.994</td>
<td>0.995</td>
<td>0.991</td>
<td>0.993</td>
<td>0.991</td>
<td>0.995</td>
<td>0.995</td>
<td>0.989</td>
</tr>
<tr>
<td>MCRTCF</td>
<td>0.995</td>
<td>0.993</td>
<td>0.995</td>
<td>0.993</td>
<td>0.993</td>
<td>0.990</td>
<td>0.994</td>
<td>0.994</td>
<td>0.987</td>
</tr>
<tr>
<td>MCRTCT</td>
<td>0.998</td>
<td>0.996</td>
<td>0.998</td>
<td>0.996</td>
<td>0.997</td>
<td>0.994</td>
<td>0.998</td>
<td>0.998</td>
<td>0.991</td>
</tr>
<tr>
<td>CVLTCR</td>
<td>0.999</td>
<td>0.993</td>
<td>1.003</td>
<td>1.013</td>
<td>1.005</td>
<td>1.011</td>
<td>1.003</td>
<td>1.006</td>
<td>0.987</td>
</tr>
<tr>
<td>CVLTCL</td>
<td>1.022</td>
<td>1.021</td>
<td>1.024</td>
<td>1.038</td>
<td>1.030</td>
<td>1.049</td>
<td>1.016</td>
<td>1.011</td>
<td>1.023</td>
</tr>
<tr>
<td>CVLTP</td>
<td>0.906</td>
<td>0.909</td>
<td>0.923</td>
<td>0.898</td>
<td>0.912</td>
<td>0.919</td>
<td>0.922</td>
<td>0.909</td>
<td>0.927</td>
</tr>
<tr>
<td>CVLTI</td>
<td>1.298</td>
<td>1.451</td>
<td>1.253</td>
<td>1.383</td>
<td>1.363</td>
<td>1.328</td>
<td>1.284</td>
<td>1.265</td>
<td>1.506*</td>
</tr>
<tr>
<td>CCC</td>
<td>1.045</td>
<td>0.050</td>
<td>1.041</td>
<td>1.033</td>
<td>1.054</td>
<td>1.053</td>
<td>1.028</td>
<td>1.025</td>
<td>0.993</td>
</tr>
</tbody>
</table>

* variation of odds ratio greater than 15% when compared to y
iii) Neuropsychological and psychosocial tests: rTMD versus IBS groups

From Table XII, the neuropsychological reaction time tests (scores 0 to infinite) (SRT, MCRT, MCRTCF and MCRTCT) showed no statistically significant difference between rTMD and IBS groups (Student's t-test and Mann-Whitney U-Wilcoxon Rank Sum Test, p > 0.05). In all tests but one (SRT), IBS patients had higher reaction time test results and were a little slower than those with rTMD, but the difference observed between the two groups in all reaction time tests was lower (range 8 - 22 msec) than our sample size calculation for 75 msec and were not considered to be relevant (Stuss et al., 1985, 1989a, 1989b).

The neuropsychological tests which evaluates attention and short-term memory under interference (CVLTCR, CVLTCL, CVLTP, CVLTI, and CCC) showed statistically significant differences between the two groups (Student's t-test, Mann-Whitney U-Wilcoxon Rank Sum Test, p<0.05)(Table III). The CVLTCR (scores 0 - 80) showed a highly statistically significant difference between the two groups (p<0.01), with IBS patients remembering fewer correct words in a shopping list (mean = 53, SD = 9) than those with rTMD (mean = 60, SD = 8.5). These were also statistically significant differences (p<0.05) between the two groups in CVLTCL scores, with IBS patients grouping less words semantically (e.g. different spices, clothes, tools and fruits) (mean = 19, SD = 10) than the rTMD group (mean = 27, SD = 11). Differences in the CVLTI and
the CVLTP were non-significant. The CCC (scores 0 - 45) also showed statistically significant difference (p<0.05) between the two groups, with IBS patients identifying fewer correct trigrams (group of three letters) (mean = 31, SD = 4.8) than those with rTMD (mean = 34, SD = 6.3). Our neuropsychological test results are extremely similar to the ones presented on Table III, with IBS patients having statistically significant worse neuropsychological tests than rTMD ones in the CVLTCR, CVLTCL and CCC.

Similar to Table III, the psychosocial tests (100 mm VAS for fatigue and energy level, Beck Depression Inventory and the Sleep Assessment Questionnaire) showed also significant results. In the University of Toronto Sleep Assessment Questionnaire (SAQ) (scores 0 - 68), IBS patients had significantly (Student’s t-test, Mann Whitney U-Wilcoxon Rank Sum Test, p < 0.01) more sleep disorders (mean = 26, SD = 8.3) than the rTMD group (mean = 20, SD = 6.2). In the BDI (scores 0 - 63), the levels of depression in the IBS group (mean = 15, SD = 9.7) was significantly higher than those patients with rTMD (mean = 7.8, SD = 6.4) (p<0.01). Fatigue (100 mm VAS) was higher for IBS (mean = 65, SD = 24) than for rTMD patients (mean = 46, SD = 27) and was also statistically different (p < 0.05). Conversely, energy level (100 mm VAS) was statistically higher for rTMD (mean = 50, SD = 24) than for IBS patients (mean = 29, SD = 19). Our psychosocial findings were also similar to the ones described for the TMD groups, with the exception that statistical differences were found for depression and energy levels.
TABLE XII: NEUROPSYCHOLOGICAL TEST RESULTS BETWEEN RESPONDING (GROUP I) AND IRRITABLE BOWEL SYNDROME (GROUP III) PATIENTS

<table>
<thead>
<tr>
<th>Independent Variables: Unit or category</th>
<th>Responding TMD, Group I (n=36) Mean (SD)</th>
<th>IBS, Group III (n=20) Mean (SD)</th>
<th>Student's t-test</th>
<th>Mann-Whitney U-Wilcoxon Rank Sum Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Reaction Time: scores (0-∞) msec</td>
<td>249 (60)</td>
<td>241 (34)</td>
<td>p = 0.5; NS</td>
<td>p = 0.7; NS</td>
</tr>
<tr>
<td>Multiple Choice Reaction Time: scores (0-∞) msec</td>
<td>437 (61)</td>
<td>452 (47)</td>
<td>p = 0.3; NS</td>
<td>p = 0.2; NS</td>
</tr>
<tr>
<td>Multiple Choice Reaction Time with Conflict: scores (0-∞) msec</td>
<td>484 (73)</td>
<td>492 (56)</td>
<td>p = 0.6; NS</td>
<td>p = 0.4; NS</td>
</tr>
<tr>
<td>Multiple Choice Reaction Time with Constraint: scores (0-∞) msec</td>
<td>447 (66)</td>
<td>469 (71)</td>
<td>p = 0.2; NS</td>
<td>p = 0.3; NS</td>
</tr>
<tr>
<td>CVLT-CR: scores 0-80</td>
<td>60 (8.5)</td>
<td>53 (9)</td>
<td>p = 0.008**</td>
<td>p = 0.007**</td>
</tr>
<tr>
<td>CVLT-CL: scores 0-60</td>
<td>27 (11)</td>
<td>19 (10)</td>
<td>p = 0.01*</td>
<td>p = 0.01*</td>
</tr>
<tr>
<td>CVLT-P: scores 0-40</td>
<td>5.3 (6.2)</td>
<td>4.3 (2.9)</td>
<td>p = 0.4; NS</td>
<td>p = 0.8; NS</td>
</tr>
<tr>
<td>CVLT-I: scores 0-10</td>
<td>0.6 (1.0)</td>
<td>1.4 (1.8)</td>
<td>p = 0.07; NS</td>
<td>p = 0.18; NS</td>
</tr>
<tr>
<td>CCC: scores 0-45</td>
<td>34 (6.3)</td>
<td>31 (4.8)</td>
<td>p = 0.04*</td>
<td>p = 0.04*</td>
</tr>
<tr>
<td>SAQ: scores 0-68</td>
<td>20 (6.2)</td>
<td>26 (8.3)</td>
<td>p = 0.005**</td>
<td>p = 0.01*</td>
</tr>
<tr>
<td>BDI: scores 0-63</td>
<td>7.8 (6.4)</td>
<td>15 (9.7)</td>
<td>p = 0.003**</td>
<td>p = 0.002**</td>
</tr>
<tr>
<td>FATIGUE: VAS 0-100 mm</td>
<td>46 (27)</td>
<td>65 (24)</td>
<td>p = 0.01*</td>
<td>p = 0.01*</td>
</tr>
<tr>
<td>ENERGY LEVEL: VAS 0-100 mm</td>
<td>50 (24)</td>
<td>29 (19)</td>
<td>p = 0.003**</td>
<td>p = 0.004**</td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01, *** p < 0.001
iv) Confounders: rTMD versus IBS

Similar to Table X, the same confounders which could not be controlled during the design stage were analyzed using the same statistical tests. In addition, clinical variables (e.g. pain intensity and pain duration) were not analyzed for the same reasons described. Recoding procedures were exactly the same as described on Table X.

On Table XIII, educational level showed no statistically significant difference between the rTMD and IBS groups, and the percentage of individuals with post-secondary educational was almost the same for both groups (74.3 versus 75%, respectively). Employment showed also no statistically significant difference between the two groups, but the p-value was closed to significance (p=0.07). The percentage of employed subjects in rTMD patients was higher than in those with IBS (60.6% versus 35.0%, respectively). Similar to employment, no statistically significant difference was found for income level, but the p-value was marginally significant (p = 0.07). The proportion of subjects with low income level for IBS patients (65%) was higher than for the rTMD group (39.4%). The average age for rTMD patients (mean = 29.4, SD = 9.0) was higher than for the IBS group (mean = 32.9, SD = 10.5) and non significant. Similar to Table X, none of the confounders may have played a role in our neuropsychological and psychosocial test results; however, employment and income may have been found significant if the sample size was higher. Logistic regression will be used below to reassess all factors.
In our logistic regression analysis, we assessed the presence of confounders by logistic regression for the same reasons described for rTMD versus nrTMD and nrTMD versus IBS groups. Similarly to Table XI, seven confounders were be included in our analysis. Similarly to described above, we considered to be a confounder any variable which changes the initial ‘y’ value in more than 15% (Greenland, 1989). From Table XIV, we can be determined that none of the seven confounders included in our logistic regression analysis affected none of the neuropsychological tests in any direction (false positives or false negatives). Similar to Table VII, the Odds Ratio for all neuropsychological tests did not change significantly with the exception of the CVLTI. In
the CVLTI, the SAQ and the BDI affected the test results. However, this finding was not considered to be relevant for the same reasons explained on Table VII. Similarly, the range of percent correct observations was also very high (73.9 to 82.6%).

Therefore, the neuropsychological tests results from Table XII were also unconfounded for age, fatigue, energy level, sleep, depression, educational level and income. From our results, it can be assessed that the neuropsychological test results between responding TMD (Group I) and irritable bowel syndrome groups (Group III) were statistically significant from each other in the following neuropsychological tests: CVLTCR, CVLTCL and CCC even when controlling for confounders. The results were extremely similar to ones described between the two TMD groups.
<table>
<thead>
<tr>
<th>Confounder</th>
<th>Age</th>
<th>Fatigue</th>
<th>Energy Level</th>
<th>SAQ</th>
<th>BDI</th>
<th>Educational level</th>
<th>Income</th>
<th>All variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRT</td>
<td>0.996</td>
<td>0.995</td>
<td>0.993</td>
<td>0.994</td>
<td>0.994</td>
<td>0.996</td>
<td>0.997</td>
<td>0.989</td>
</tr>
<tr>
<td>MCRT</td>
<td>1.005</td>
<td>1.006</td>
<td>1.001</td>
<td>1.000</td>
<td>0.997</td>
<td>0.998</td>
<td>0.995</td>
<td>0.990</td>
</tr>
<tr>
<td>MCRTCF</td>
<td>1.001</td>
<td>1.001</td>
<td>0.999</td>
<td>0.996</td>
<td>0.993</td>
<td>0.994</td>
<td>1.001</td>
<td>0.999</td>
</tr>
<tr>
<td>MCRTCT</td>
<td>1.004</td>
<td>1.005</td>
<td>1.003</td>
<td>1.000</td>
<td>0.998</td>
<td>0.998</td>
<td>1.005</td>
<td>1.004</td>
</tr>
<tr>
<td>CVLTCR</td>
<td>0.920</td>
<td>0.918</td>
<td>0.941</td>
<td>0.936</td>
<td>0.946</td>
<td>0.931</td>
<td>0.920</td>
<td>0.935</td>
</tr>
<tr>
<td>CVLTCL</td>
<td>0.930</td>
<td>0.913</td>
<td>0.945</td>
<td>0.942</td>
<td>0.946</td>
<td>0.936</td>
<td>0.927</td>
<td>0.943</td>
</tr>
<tr>
<td>CVLTP</td>
<td>0.960</td>
<td>0.963</td>
<td>0.950</td>
<td>0.984</td>
<td>0.998</td>
<td>1.005</td>
<td>0.957</td>
<td>0.986</td>
</tr>
<tr>
<td>CVLTI</td>
<td>1.531</td>
<td>1.455</td>
<td>1.504</td>
<td>1.610</td>
<td>1.791</td>
<td>1.771</td>
<td>1.555</td>
<td>1.529</td>
</tr>
<tr>
<td>CTT</td>
<td>0.902</td>
<td>0.898</td>
<td>0.881</td>
<td>0.878</td>
<td>0.917</td>
<td>0.918</td>
<td>0.876</td>
<td>0.873</td>
</tr>
</tbody>
</table>

* variation of odds ratio greater than 15% when compared to y
g) Associations between TMD and IBS patients versus the non-pain group

We have obtained data on 36 rTMD, 24 nrTMD, 20 IBS patients as well as on 15 language-, age- and sex-matched non-pain subjects. Five subjects of the non-pain group, initial size was 20, had to be excluded due to very high scores in the Sleep Assessment Questionnaire, which was one of our criteria for exclusion. The variables described were only the ones tested for all groups; therefore, the clinical variables which were not tested for both IBS and non-pain groups were not included. One-way analysis of variance (ANOVA, p < 0.05), which detects differences in means among three or more means, and the Tukey-b Multiple Range test (p < 0.05), which detects differences in means between groups, were used to assess all four groups at the same time in order to compare our patient versus our non-pain populations. The results were not assessed for confounders, because this was neither both of our hypothesis.

Table XV shows statistically significant differences in the following tests: the CVLTCR (p < 0.01), the CVLTCL (p < 0.01), the CCC (p < 0.001), fatigue (p < 0.001), energy level (p < 0.01), the SAQ (p < 0.0001), and the BDI (p<0.001). This findings are similar to the ones already described in the 2X2 tables and demonstrates that the four groups most likely did not come from the same population. In addition, no statistically significant differences could be found for any of the reaction time tests.

Our most important findings come from the differences between the four groups rather than among. Multiple Range (Tukey-b) test was used to detect differences between groups. The differences between rTMD versus nrTMD, nrTMD versus IBS, and rTMD
versus IBS have already been described. We concentrated our analysis in the comparisons between the non-pain group (Group IV) versus the TMD and IBS groups.

The language-, age- and sex-matched non-pain comparison group showed no statistically significant difference with the other patient groups in all reaction time tests. All test results were comparable, but the non-pain group tended to be a little faster than the other groups, but the differences range between (1 to 80 msec) which is higher than our sample size estimate (75 msec) but considered to be non-relevant (Stuss et al., 1985, 1989a, 1989b). Statistically significant differences (Multiple Range Tukey-b test, p<0.05) were found between the non-pain and IBS groups in the following tests: a) CVLTCR, b) CCC, c) fatigue, d) energy level, e) SAQ, and f) BDI. The non-pain group showed better neuropsychological test results, higher energy levels, lower levels of fatigue, lower scores in the SAQ, and lower BDI scores than the IBS group.

The non-pain group also had statistically significant differences, similar to observed with IBS patients, when compared to nrTMD in the following tests: a) CVLTCR, b) CCC, c) fatigue, d) SAQ, and e) BDI. Analogously to IBS, the non-pain group showed better neuropsychological test results, higher energy levels, lower levels of fatigue, lower scores in the SAQ, and lower BDI scores than the nrTMD group.

When comparing to rTMD patients, only one test was statistically significant, the SAQ, where the sleep scores for the rTMD group (mean=20, SD=6.2) was higher than the non-pain group (mean=11, SD=4.7). However, despite non-significant, non-pain subjects also showed neuropsychological test results, fatigue and energy levels, and depression scores better than patients with rTMD.
Our findings suggest that the neuropsychological scores from a non-pain population are more similar to the rTMD group than to both the nrTMD and the IBS groups. Finally, it should also be stressed that these findings were not controlled for confounders and must be used only as hypothesis generating.
<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Responding TMD, Group I (n=36)</th>
<th>Non-responding TMD, Group II (n=24)</th>
<th>IBS Group II (n=20)</th>
<th>Normal Controls Group IV (n=15)</th>
<th>One-way ANOVA (P &lt; 0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td>SRT: (0-∞) msec</td>
<td>249 (60)</td>
<td>261 (67)</td>
<td>241 (34)</td>
<td>223 (24)</td>
<td>p = 0.20; NS</td>
</tr>
<tr>
<td>MCRT: (0-∞) msec</td>
<td>437 (61)</td>
<td>477 (92)</td>
<td>452 (47)</td>
<td>437 (69)</td>
<td>p = 0.14; NS</td>
</tr>
<tr>
<td>MCRTCF: (0-∞) msec</td>
<td>484 (73)</td>
<td>528 (107)</td>
<td>492 (56)</td>
<td>467 (57)</td>
<td>p = 0.09; NS</td>
</tr>
<tr>
<td>MCRTCT: (0-∞) msec</td>
<td>447 (66)</td>
<td>480 (99)</td>
<td>469 (71)</td>
<td>433 (50)</td>
<td>p = 0.19; NS</td>
</tr>
<tr>
<td>CVLT-CR: scores 0-80</td>
<td>60 (8.5)</td>
<td>53 (10) a</td>
<td>53 (9) b</td>
<td>62 (8.9) d, e</td>
<td>p = 0.001**</td>
</tr>
<tr>
<td>CVLT-CL: scores 0-60</td>
<td>27 (11)</td>
<td>18 (7.4) a</td>
<td>19 (10) b</td>
<td>25 (13)</td>
<td>p = 0.003**</td>
</tr>
<tr>
<td>CVLT-P: scores 0-40</td>
<td>5.3 (6.2)</td>
<td>5.6 (4.2) a</td>
<td>4.3 (2.9)</td>
<td>2.6 (2.7)</td>
<td>p = 0.20; NS</td>
</tr>
<tr>
<td>CVLT-I: scores 0-10</td>
<td>0.6 (1.0)</td>
<td>0.7 (1.4) a</td>
<td>1.4 (1.8)</td>
<td>0.7 (0.8)</td>
<td>p = 0.14; NS</td>
</tr>
<tr>
<td>CCC: scores 0-45</td>
<td>34 (6.3)</td>
<td>30 (6.3) a</td>
<td>31 (4.8)</td>
<td>38 (3.8) d, e</td>
<td>p = 0.0001***</td>
</tr>
<tr>
<td>FATIGUE: VAS 0-100 mm</td>
<td>46 (27)</td>
<td>68 (25) a</td>
<td>65 (24) b</td>
<td>39 (21) d, e</td>
<td>p = 0.0007***</td>
</tr>
<tr>
<td>ENERGY: VAS 0-100 mm</td>
<td>50 (24)</td>
<td>44 (25)</td>
<td>29 (19) b</td>
<td>55 (11) e</td>
<td>p = 0.004**</td>
</tr>
<tr>
<td>SAQ: scores 0-68</td>
<td>20 (6.2)</td>
<td>24 (6.8) a</td>
<td>26 (8.3) b</td>
<td>11 (4.7) c, d, e</td>
<td>p = 0.0000***</td>
</tr>
<tr>
<td>BDI: scores 0-63</td>
<td>7.8 (6.4)</td>
<td>11 (6.9)</td>
<td>15 (9.7) b</td>
<td>2.8 (3.2) d, e</td>
<td>p = 0.0000***</td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01, *** p < 0.001
a difference in means between rTMD (Group I) and nrTMD (Group II), b difference in means of rTMD and IBS, e difference in means between non-pain population and rTMD patients, d difference in means between non-pain population and nrTMD patients, e difference in means between non-pain population and IBS patients (Tukey-b Multiple Range test, p < 0.05)
VII. DISCUSSION

a) Sample size, primary diagnosis, recruitment rate and follow-up

In this investigation, sample size was a little smaller than other studies (n=60) for the TMD group as compared to Fricton & Olsen's (n = 94), Schnurr and coworkers' (n=178), Millstein-Prentky & Olson's (n = 135), but it was similar to Lipton & Marbach's (n = 68) and larger than Gerschman and colleagues' (n = 47). The sample size was strictly calculated based on the smallest difference in reaction time tests from a previous study (Goldberg et al., 1996). Because of that, it was not possible to include a great number of variables (n = 45), and only the difference in neuropsychological tests and their specific confounders was assessed.

Similar to other studies (Fricton & Olsen, 1996; Schnurr et al., 1991; Gerschman et al., 1987; Lipton & Marbach, 1984), the primary diagnosis was myofascial pain with or without TMD disc displacement. This investigation was very restrictive and only 19% of the new patients presented for orofacial pain treatment met the inclusion/exclusion criteria. This is in agreement with a current trend in the TMD literature to study well defined populations in order to increase the reproducibility of the results (Dworkin & LeResche, 1992). One interesting finding was that 81% of patients seeking treatment for orofacial pain were not included in the TMD diagnosis. The remaining diagnoses; e.g. post-traumatic TMD due to Motor Vehicle Accidents (MVA) (10%), neurological
disorders (9%), headaches (8%), TMJ pain without muscle involvement (7%), atypical facial pain (7%), psychological disorders (7%), and oral mucosal diseases (6%); have completely different etiologies, treatment planning and success rates than TMD. For example, patients with post-traumatic TMD as a result of a MVA have a success rate significantly reduced as compared to the non-traumatic or idiopathic TMD group (48% versus 80%, respectively) and higher proportion of affective disorders (Romanelli, Mock & Tenenbaum, 1992). The inclusion of one or more of these groups in our analysis probably would have changed our results and conclusions substantially (Appendix I).

The recruitment rate was moderate (50%), but lower than other studies. In Fricton & Olsen's study (1996), the recruitment rate was better (90%); but their rate at follow-up was 76% which is worse than in the present study (100%). In the study by Schnurr and colleagues (1991), one hundred patients did not respond to the follow-up questionnaire (43.8%). However, despite this discrepancy of drop-out rate among studies, this does not appear to affect treatment success, because the percentage of replying patients who believed that treatment was successful (31.4%) was approximately equal to those who thought that it was not successful (34.8%) (Schnurr et al., 1991).

b) Sociodemographic description of the populations studied at baseline

Regarding gender, all subjects were females in order to increase our internal validity and to control an important confounder in the design stage (Hennekens & Buring,
1987). Despite that, the gender distribution seems to be comparable to other studies considering that the percentage of women has been much greater than men. In Fricton & Olsen's study (1996), 89% were females; in Gerschman et al. (1987), 74.9%; in Lipton & Marbach's, 76.5%; Strychalski et al., 65%; in Millstein-Prentky & Olson's, 66%; and in Schwartz and colleagues, 100%.

The majority of the TMD population (64.4%) had post-secondary diploma certificate or higher. This is in agreement with previous studies which also described high educational levels for the TMD population. Fricton & Olsen (1996) reported that the average education for the TMD population was 2 years of college. Gerschman et al. (1987) described similar findings with and educational average of 10 years. In the same way, Lipton & Marbach (1984) also found high levels of education in the TMD group (mean = 13.1 years, SD = 2.8).

The TMD population also had high levels of employment (67.8%), but it has predominantly low levels of income (57.9%). This contrasts with a previous study by Gerschman et al. (1987) in which 73.6 % of the TMD patients had good financial situation. Unfortunately, not all studies reported employment and income levels and it was difficult to assess their overall differences.

The average age of the TMD sample was 28.3 (SD = 9.0, range 12 - 46). Only one patient was under 15 years, which was one of our inclusion criteria, but it was included in the study for reasons already described in the results section. The results in this investigation are also similar to previous studies. In Fricton & Olsen's investigation, the age of the TMD patients ranged from 16 to 61 years (mean = 36.6 years, SD = 11.7).
Schnurr et al. (1991) reported a mean age of 27.4 years (SD = 9.7). Gerschman et al. (1987) found a higher range and average age (mean = 47, range 5 - 60 years). With the exception Gerschman and colleagues' work, in all studies the overall average age was around 30 years with a standard deviation of 10.

When compared to the other two groups (irritable bowel syndrome and non-pain population), no statistical significant difference was found regarding educational level, employment, income and age. The values for the irritable bowel syndrome are comparable to the ones described by Thompson (1992) and Toner (1994) for age. In the overall, no statistically significant differences were observed in sociodemographic variables used in this study (e.g. gender, educational level, employment, income and age) among the three groups studied at baseline. In addition, the demographic description of the TMD population seems to be comparable with other TMD treatment outcome studies.

c) Pain improvement

At six-month follow-up, the TMD group (n = 60) was assessed for pain improvement after treatment. This study employed two improvement criteria: a) a non-verbal, thirty percent reduction or greater in 100 mm VAS in pain at rest over baseline, and b) a verbal one, patient self-assessment (better, same or worse) (Table II). The improvement rate in the non-verbal scale was 60% (36 out of 60 cases), while in the verbal one was 63% (38 out of 60). The overall percent agreement was 90% and the agreement controlling for chance (Kappa) was 0.78 which is considered substantial by
Landis & Koch (1977). Although the improvement rate in this study is not as large as the ones described in the literature (70% to 90% success rate), this may have been a result of the very restrictive inclusion criteria used which contrasts with previous studies where a inclusion/exclusion criteria was not reported (Greene & Laskin, 1983).

The results in this investigation are very similar to ones described by Fricton & Olsen (1996). They assessed their TMD patients also at six-month and one year after baseline. Their improvement criteria was a decrease of one SD (0.14) or greater on the Craniomandibular Index (CMI) and if post-treatment CMI scores were less than 0.15. In addition, they also used another index for improvement: the Symptom Severity Index (SSI). A 30% decrease in SSI ratings was considered as an acceptable level of change in subjective pain severity, since this percentage is equivalent to one standard deviation when employing pre-treatment SSI values (Fricton & Olsen, 1996). In their study, 29 of 47 subjects (61.7%) in the criterion group and 27 of the 47 subjects (57.4%) in the cross-validation group were classified as successfully treated at post-treatment. Lipton & Marbach (1984) found a success rate after 5 month follow-up of 57% measuring reduction in the level of psychological distress (Langner 22-item index Symptom). Millstein-Prentky & Olson (1979) found 62% improvement on the Minnesota Multiphasic Personality Inventory (MMPI). These findings are important, because they were replicated in a study with similar population description and time for follow-up with almost the same improvement rate. The only difference was that in Fricton & Olsen’s study, a combination of indexes were used instead of a unidimensional evaluation as in
this one. Therefore, unidimensional very multidimensional evaluations seems to provide similar improvement rates in studies with similar methodologies.

However, it must be kept in mind that there is no perfect means of assessment, and all instruments have their advantages and disadvantages. The selection of the measuring instrument must be determined based on for whom the study is designed for (e.g. patient, insurance companies or health care units) and the type of population being studied (high versus low level of education and income, language skills, etc.). A population with low levels of education and/or poor language skills may prefer a simple form of pain assessment (Magnusson et al., 1995). In addition, the pain experience has sensory, affective and cognitive components and it is difficult to be interpreted. This study agrees with previous ones which found that different verbal and non-verbal scales correlate well with each other (Carlsson, 1983; Magnusson et al., 1995).

This investigation, however, disagrees in part with the previous ones which affirm that caution must be taken when assessing pain improvement by reduction in baseline assessment using VAS; because, pain memory is less precise for chronic pain patients than those with acute or experimental pain. Despite being true that chronic pain patients usually remember initial pain intensity at follow-up higher than observed at baseline (Carlsson, 1983), this per se does not invalidate the VAS as a treatment outcome measure. The variation that has been reported for pain memory does not exceed 23% improvement (Erskine et al., 1990). Therefore, when assessing improvement, this variation in response must be taken into account besides the variability in the measurement instrument itself. In this study, the non-verbal improvement criteria of 30%
in pain at rest over baseline is greater than the reported pain measurement error (plus or minus 5 mm) and also the 23% variability in pain memory (Carlsson, 1983; Price et al., 1994; Magnusson et al., 1995). Finally, the average improvement rate for the responding TMD group was 66%, and it is very difficult to attribute the high improvement rate found in this study only to variation in pain assessment and to the variation in pain memory itself.

In addition, the results in this study also correlate very well with the patient’s self-assessment of pain improvement (Kappa = 0.78). These reported findings are similar to Colvin et al. (1980) who reported that 97% of the patients (n = 287) indicated that they would accept a 50% improvement and 52% would try elsewhere if the treatment did not bring acceptable pain relief. The only difference in this investigation is the fact that 30%, rather than 50%, reduction was considered good improvement by 94% of our patients (Table II) who considered that the treatment was beneficial. The patient satisfaction with the treatment is important, because Cassisi et al. (1989) reported that patients who completed treatment and improved (self-assessment) reported significantly fewer physician visits, fewer surgeries following treatment, higher return rates to the working place, than those who were not approved for treatment by insurance carriers or those who declined treatment by themselves. The patient satisfaction with the treatment can readily be converted to dollars saved for insurance companies and the health care system. In addition, the selection of an appropriate cutoff point also prevents the repetitive use of different pain scales. This might influence the correlation between the two types of
estimates of pain intensity; because, the patient may learn the use of the scales, which might result in a higher correlation and bias the results (Carlsson, 1983).

Considering the importance of change, we must also compare the clinical versus the statistical significance of treatment outcome. Regarding treatment efficacy, we must also compare post-treatment levels with pre-treatment levels, but also with non-pain individuals (Jacobson et al., 1984). In this research, the average post-treatment pain levels for the responding TMD (Table I) was still considerably higher than the non-pain group (21 mm versus 4 mm on 100 mm VAS, respectively). This also exemplifies that the treatment goal should be pain reduction and not pain elimination.

Temporomandibular disorder patients were not followed after 6 months; however, the literature suggests that the long term success is also high. In Heloe and colleagues’ study (1980), the improvement after 1 1/2 year was 81% of patients reporting improvement or partial improvement. In a study by Strychalski et al. (1984) which follows TMD patients for periods ranging from 2-3 years after treatment, still found 72% improvement. The treatment results seem to be stable even for longer periods ranging from 1 to 9 years after treatment, 60 to 98% (Greene & Laskin, 1983).

d) Predictors of treatment outcome for TMD patients

Previous studies comparing patients with recovered mild concussion (Closed Head Injury - CHI) with control subjects were capable of differentiating them in terms of performance in neuropsychological tests (Stuss et al., 1985). Analogously, a study by
Goldberg et al. (1996) was also capable of differentiating between idiopathic and post-traumatic TMD groups using the same tests of the previous author.

Similarly, using similar methodology, the results in this investigation also suggest that different chronic pain populations, in our case rTMD versus nrTMD groups, might be separated from one another based on neuropsychological tests. On average, the nrTMD population performed worse in the neuropsychological tests than the rTMD in the memory tests, but no difference was found in the reaction time tests. The overall interpretation of the results in the neuropsychological tests, psychosocial tests, confounders, signs and symptoms of TMD, occlusal variables and pain intensity will be further discussed below. Most of the discussion will be in light of the studies which used similar neuropsychological tests rather than the general literature which used different neuropsychological tests in different populations. This was done, because it is very difficult to draw conclusions from studies with both diverse methodologies and populations.

i) Neuropsychological tests

The reaction time test results in this study were very similar to ones published by Goldberg et al. (1996); Stuss et al. (1989a, 1989b) for the non-traumatic TMD and non-pain populations. The reaction time tests had the following range of results of both rTMD and nrTMD groups: a) SRT, range 249 - 261 milliseconds; b) MCRT, range 437 to 477
milliseconds; c) MCRTCF, 484 to 528 milliseconds; and d) MCRTCT, 447 to 480 (Table III). This was comparable to Goldberg and colleagues’ study which found the following results for non-traumatic TMD population: a) SRT, 304 milliseconds; b) MCRT, 486 milliseconds; and c) MCRTCF 529 milliseconds. According to Stuss et al. (1989a, 1989b), differences smaller than 100 milliseconds are not clinically relevant. The maximum difference in my findings as compared to the above mentioned author ranged from 45 to 55 milliseconds, which was far below one hundred. In Stuss et al. (1989a, 1989b), the results of the neuropsychological tests for their control groups ranged from: a) SRT, 235 - 265 milliseconds; b) MCRT, 430 - 433; c) MCRTCF, 512 - 525; and d) MCRTCT, 435 - 439. Comparing to Stuss and coworker’s results, the test differences in this investigation ranged from 7 to 45 milliseconds, which was also considerably below than 100 milliseconds. Finally, Stuss et al. (1989b) have demonstrated that the reaction time tests were not significantly affected both by age and gender, unless the patient was over 60 years old, which is not in our inclusion criteria for age (15 to 45 years).

The reaction time tests in this study, on the other hand, contrasted with the ones by Goldberg and coworkers’ study, because differences between the two comparison groups were not found (rTMD and nrTMD) as they were by previous authors, comparing post-traumatic versus non-traumatic TMD patients in all reaction time tests. The test results between my rTMD and nrTMD groups also differed substantially from those with post-traumatic TMD. Interestingly, the test results of the post-traumatic TMD (as a result of a Motor Vehicle Accident) were comparable to those of Stuss et al. (1989a) for patients with non-hospitalized mildly concussed patients (traumatic brain injury - TBI).
Taking also this study as a comparison, these findings suggest that the patients with post-traumatic TMD might have a similar etiology to those with TBI. In addition, reaction time tests could be used to identify those patients with traumatic TMJ without brain injury from those with brain involvement. This is very important, because the success rate between traumatic versus idiopathic TMD differs significantly (48% versus 80%, respectively) (Romanelli, Mock & Tenenbaum, 1992). However, this should be addressed in a specific study.

In all reaction time tests, nrTMD was slower than rTMD, but the actual differences observed between the two groups in all reaction time tests was very small (range 12 - 44 msec) which did not reach the minimal difference determined by the sample size calculation (75 msec) and was considered non-relevant according to Stuss et al. (1985, 1989a, 1989b).

Some of the neuropsychological tests employed in this study which evaluate attention, short-term and short-term memory under interference (CVLTCR, CVLTCL, CVLTP, CVLTI, and the CCC) showed statistically significant differences between the rTMD versus nrTMD populations (Student’s t-test, Mann-Whitney U-Wilcoxon Rank Sum Test, p<0.05)(Table III). The CVLTCR, which measures attention and short-term verbal memory, showed a statistically significant difference between the two groups (p<0.01), with nrTMD patients remembering less correct words in a shopping list (mean = 53, SD = 10) than Group I (responding TMD) (mean = 60, SD = 8.5). My results in the CVLTCR (mean range 53 - 60) for both TMD groups were comparable to the ones reported by Goldberg et al. (1996) in which the results for the idiopathic TMD had an
average of 55.6 (SD = 3.18). In addition, the results in this study were also substantially higher than the post-traumatic TMD (mean 44.7, SD = 3.06), which in Goldberg and colleague’s have shown to be statistically different from the idiopathic TMD group (p < 0.05). Additionally, our non-pain population value (Table XV) is also comparable to the test external normal values (Delis et al., 1987). In the CVLT manual, the mean value ranged from 62 - 64 (T50); while in this investigation was 62 (SD = 8.9), demonstrating the reliability of the test with language-, age- (17 - 44 years) and sex-matched (women only) populations.

In a similar way, the CVLTCL, which measures not only attention but also verbal cognition, also showed statistically significant difference (p<0.001) between the two groups, with nrTMD patients grouping less words semantically (e.g. different spices, clothes, tools and fruits) (mean = 18, SD = 7.4) than rTMD ones (mean = 27, SD = 11). The results in the CVLTCL for both TMD groups were also comparable to the ones reported by Goldberg et al. (1996) for idiopathic TMD patients (mean = 21, SD = 3.9). Additionally, the results in this study were also substantially higher than the post-traumatic TMD (mean 12, SD = 2.1), which in Goldberg and coworker’s have shown to be not only higher, but also statistically different from the idiopathic TMD group (p < 0.05) showing reproducibility of the test results. The non-pain population had a mean of 25 (SD = 13); however, considering that the CVLTCL was a sub-section of the CVLT, no standard norms were available (Delis et al., 1987). Analogously, the CVLTI (number of intrusion of words which did not belong to the list) and the CVLTP (perseverations or repetitions of words in the list) had no external control values to be compared with the
non-pain population (means = 2.6 and 0.7, respectively); notwithstanding, their absolute values were very low. Additionally, the differences between the rTMD versus nrTMD groups were non-significant for both tests. Goldberg et al. (1996) found lower results in the CVLTP for the idiopathic TMD groups (mean = 3.6, SD = 0.72) when compared to our TMD population. Their results did not differ significantly from the post-traumatic TMD population (mean = 5.5, SD = 1.0) which also did not differ from both the rTMD and nrTMD groups in this investigation (means = 5.3 and 5.6, respectively). One of the possible explanations why the results in this study for the CVLTP were more similar to the post-traumatic TMD group (pTMD) than the idiopathic TMD (iTMD) is the fact the values were low, which increases the variability of the results. Another evidence that may support that assumption is the fact that in both studies, the CVLTP values were not significantly different than their comparison groups. Finally, Goldberg et al. (1996) did not use the CVLTI which does not allow us comparisons of the values of this study with both the iTMD and pTMD.

The Brown-Peterson Consonant Trigram Auditory Memory Task (CCC), which measures short-term memory under interference, also showed statistically significant difference (p<0.01) between nrTMD and rTMD, with nrTMD scoring less correct trigrams (group of three letters) (mean = 30, SD = 6.3) than rTMD (mean = 34, SD = 6.3). Our internal control values (non-pain group) (mean = 38.1, SD = 3.8) were comparable to reported age- (20 - 40 years) and sex-matched external controls (mean = 34.4 - 35, SD = 2.2 - 2.6). This confirms Stuss et al. (1987) findings that the CCC was not very influenced by both age and gender differences.
Similar to what was observed with the reaction time tests, the test results in this study for the CVLTCR, CVLTCL and CCC were similar to the idiopathic TMD group, while the post-traumatic TMD group seemed to be more similar to those with traumatic brain injury (TBI). The difference is that now the memory tests can also be used as predictors between the rTMD versus the nrTMD population. The reason(s) why nrTMD patients scored lower in the neuropsychological (memory) tests but not in the neuropsychological (reaction time) tests is (are) still unknown, but possible explanations will be further discussed below in light of the psychosocial results.

ii) Psychosocial variables

Fatigue, energy level, the Beck Depression Inventory and the Sleep Assessment Questionnaire showed that nrTMD had higher fatigue and lower energy levels when compared to rTMD patients (Table III). In the same way, nrTMD patients showed greater levels of depression and sleep disorders than rTMD ones. However, not all tests were good predictors. The SAQ demonstrated to be a good predictor, having nrTMD patients significantly (p < 0.05) more sleep disorders (mean = 24, SD = 6.8) than rTMD ones (mean = 20, SD = 6.2). The non-pain population values (mean = 11.7, SD = 4.7) were comparable to 30 external controls published by Cesta, Moldofsky & Sammut (1996) (mean = 10.8, SD = 5.7). In addition, the sleep scores range from both TMD and IBS populations (20 - 26) was comparable to the patients with primary diagnoses of sleep apnea, periodic leg movements, and snoring (mean = 26.0, SD = 8.6). The proportion of
TMD individuals with sleep disorders (cutoff = 16, sensitivity = 73, specificity = 80) reached 78% of individuals with sleep disorders, with rTMD patients having 68.6% and nrTMD, 91.7%. The non-pain population showed only 27% of individuals classified with sleep disorders. Fricton & Olsen (1996), using stepwise multiple regression and discriminate analysis for detection of predictors of treatment outcome also found sleep habits (r = -0.22) to be a good predictor. Stepwise of the 10 best predictors also included level of sleep as one the best predictors. In addition, different than the other predictors, sleep scores have shown not only to differentiate among different chronic pain populations, but also between chronic pain and non-pain groups. This happened, because the scores of Groups I, II and III were statistically significantly different than the non-pain population (Tukey-b Multiple Range Test, p < 0.05).

On the other hand, the Beck Depression Inventory was not a good predictor, and despite the fact that the levels of depression in the nrTMD group (mean = 11, SD = 6.9) was higher than the rTMD one (mean = 7.8, SD = 6.4), the difference was not significant. This is in disagreement with the findings of Dworkin et al. (1989, 1991), where dysfunctional (non-responding) TMD patients scored significantly higher, in the top 15% to 25% of scores, on measures of depression and somatization. It must be pointed out that in this study, the p-value was marginally significant (p = 0.08), and the limited sample size might have being the reason why the test was not significant. In the rTMD group, 11.8% of the patients had scores equal or greater 16 (moderately depressed) and compared to 29.2% in the nrTMD. The overall TMD population had 18.3 % of cases classified as depressed which almost the same as the 18% described by Gerschman et al.
(1987) using the Hamilton Depression Scale. This is substantially higher than the 6% found in the non-pain population (Nielsen & Williams, 1980); however, in this investigation, the fact that no non-pain subjects were included in the depressed group might have been attributed to the small sample size. This study was also similar with the one by Schwartz et al. (1979) where they compared the MMPI profiles of 42 successfully treated and 42 unsuccessfully treated female patients with MPD syndrome. Both groups differ in terms of proportion of affective disorders marked by higher levels of depression and somatization.

One interesting observation was the fact that fatigue and energy level, despite correlated, were not perfect corollaries. Levels of fatigue (100 mm VAS) were higher for nrTMD (mean = 68, SD = 25) than for rTMD patients (mean = 46, SD = 27), and with a highly statistically significant difference (p < 0.01) between the two. On the other hand, energy level (100 mm VAS), despite higher for rTMD patients (mean = 50, SD = 24) than for nrTMD (mean = 44, SD = 26), was non-significant and showed to be a worse predictor than fatigue. In Goldberg et al. (1996), the authors reported high levels of affective disorders using the Symptom Checklist-90 Revised in both post-traumatic (54%) and idiopathic TMD (43%) populations. This investigation used sleep, depression, fatigue and energy level as confounders for the neuropsychological tests used, and the sample size was not calculated to use them as predictors. Therefore, specific studies must be carried out to check the hypothesis if these variables are still good predictors with the appropriate sample size and controlling for specific confounders. All these four variables seem to be correlated at different levels. Sleep correlated well with fatigue (r = 0.43), but
not so well with energy level \( r = -0.04 \) and depression \( r = 0.10 \). Depression correlated equally well with fatigue \( r = 0.29 \) and energy level \( r = -0.28 \). And fatigue and energy level correlated very well with one another \( r = -0.53 \) (Appendix IV). It must be kept in mind that bivariate associations are not conclusive, and cause and effect relationships cannot be established from correlations alone. Therefore, these findings can only be used for generating hypotheses (Hennekens & Buring, 1987).

The literature has shown that patients with closed head injury have both reaction time and memory tests (CCC) affected when comparing to the matched non-patient population (Stuss et al., 1985). Although there is the possibility that post-traumatic TMD and closed head injury (CHI) patients might share similar etiology, the results reported in the literature only allow comparisons in light of previous findings (Goldberg et al., 1996; Stuss et al., 1985, 1989). In addition, it has also been shown that affective disorders seems to be correlated with performance in both reaction time and memory tests (Romanelli, Mock & Tenenbaum, 1992; Goldberg et al., 1996). One of the possible explanations for the fact that non-responding TMD patients differ only in the memory tests and psychosocial variables and not in the reaction time tests, might be that higher levels of affective disorders may interfere with more complex (memory) abilities, but not with more simple functions such as reaction time tests (attention). On the other hand, traumatic injuries which affect the brain, as seen in patients with closed head injury or those patients following Motor Vehicle Accidents (post-traumatic TMD), have been shown to affect more basic functions as reaction time tests (Stuss et al., 1985, 1989;
Goldberg et al., 1996). However, these two hypotheses are still highly speculative at this time.

Different authors seem to disagree regarding the role of personality differences in the etiology of TMD. In a study by Small & Hill (1974), fifty patients (26 females, age range 14–54; 24 males, age range 19 to 47) with temporomandibular joint pain-dysfunction syndrome were interviewed and given the Minnesota Multiphasic Personality Inventory (MMPI) and Cornell Medical Index (CMI). According to these instruments, ten patients were classified as normal and 40 patients were classified as abnormal. These results indicate that specific personality types could be correlated with TMD treatment outcome. On the other hand, Millstein-Prentky & Olson (1979) reported a decline in the discriminating ability of the MMPI was due to absence of consistent differences in personality between successful and unsuccessful MPD patients. Both groups showed similar profile configurations (psychosomatic-V); however, the unsuccessful group had higher profile elevations which agrees with Schwartz et al. (1979). The results suggested that a single scale to predict treatment outcome would be inefficient due to the absence of consistent personality differences in MPD patients.

Other psychological variables, such as depression, has also been attributed to the etiology and perpetuation of TMD. In Fricton and Olsen’s study (1996), six (level of eating habits, level of sleep, level of sexual activity, low energy, low self-esteem, feeling confused) of the ten predictors correlated with depression. Gerschman et al. (1987) reported that psychiatric evaluation showed that many of these patients suffered from a chronic psychiatric illness of moderate severity with the most common diagnosis was
some variety of depression. According to the same author, the treatment outcome failure is correlated with depression in the patients who did not respond to treatment. Similarly, Greene et al. (1982) have reported that some of the complaints of MPD patients are weight loss, difficulty in eating, difficulty in sleeping, and a reduction in social activity. But these symptoms may also be the result of chronic pain. Treatments for depression alleviates pain, but a number of treatment modalities also have the same effect. In addition, similar depression prevalence between TMD and other chronic pain syndromes is not a confirmation of etiology, because the chronic pains may be the initiating factor, and depression one of the consequences. The implications of these findings for the treatment of TMD pain and chronic pain in general is that symptoms of depression must be addressed if treatment is to be successful. Interventions focused entirely on the pain or on physical symptoms may be misdirected. Finally, depression is not an isolated entity and it is often associated with symptoms of stress and anxiety, and stress-related oral behaviors are the presumed cause of temporomandibular disorders (Fricton & Olsen, 1996).

Other psychological variables, such a coping style, has been associated with TMD treatment outcome. According to Schnurr et al. (1991), of the five variable sets entered into the discriminant function analyses to predict changes in pain intensity, the only analysis to obtain significance was that for the Ways of Coping Scale.

After many years of studying the psychological traits and states of patients and their relationship to the etiology of TMD, it can be concluded that it is not possible to confirm the etiology of TMD by assessing past or current psychological factors for an
individual patient. However, the evidence is abundant and clear that psychological factors do have an important role in the etiology, progression and TMD treatment (Schwartz et al., 1979). Despite our understanding of the role of psychosocial factors in the etiology of temporomandibular pain and pain in general, little is yet known about the factors contributing to the perpetuation of chronic pain (Fricton & Olsen, 1996). Previous research has not emphasized the importance of psychosocial factors in predicting outcome for patients with chronic pain. Many studies have established that a substantial proportion of patients with chronic orofacial pain demonstrated psychiatric symptoms. While there is controversy about the causal relationship of pain and psychiatric symptoms, there is a lack of studies of the role of psychosocial variables with the perpetuation of chronic pain (Gerschman et al., 1987).

These described findings have implications for therapists who wish to understand the disorder and to manage patients appropriately. Clinicians should recognize the importance of nonspecific factors such as placebo effects, doctor-patient interactions, and spontaneous recoveries in the treatment response. This awareness will enhance their effectiveness as therapist and will help them to avoid excessive or radical treatment methods in favor of multidisciplinary teams for chronic pain management using reversible therapies. Our results further confirm with more sensitive methods (neuropsychological tests) that the biopsychosocial model seems to be most adequate for the understanding of temporomandibular disorders in particular and chronic pain in general.
iii) Confounders

Social variables have also been correlated with TMD treatment outcome. In a study by Gerschman et al. (1987), one-hundred and thirty patients with dental phobias and 368 patients with chronic orofacial pain were compared for psychological and social variables. Pain patients showed a greater burden of psychiatric disorders and were more likely to be older, married, have children, be migrants, be less educated, have poorer jobs and be more financially disadvantaged than phobic patients. Social and cultural factors have been found to influence the meaning of pain, reaction to pain and communication of pain as well as psychophysiological and autonomic functioning.

In this study, the objective, similar to the psychological variables, was to evaluate specific social factors that has been described as confounders for the neuropsychological tests used (Goldberg et al., 1996) and not as predictors like previous studies. Nine confounders which could not be eliminated during the design stage were included in the analysis (i.e, educational level, employment, income, age, length of treatment, pain duration, number of treatments after initial visit, improvement according to the treating clinician and comorbidity with irritable bowel syndrome), and only five were found to be either significant or relevant confounders: educational level, income, length of treatment, comorbidity with IBS, and number of treatments provided. All categorical variables were dichotomized (educational level, employment, income and comorbidity with irritable bowel syndrome) in order to calculate odds ratio (Table IV).
Comparisons with previous studies have been already described in the discussion of Table I for educational level, employment, income and age. We will now discuss the implications of our findings for length of treatment, pain duration, number of treatments provided, success rate for different treating clinicians and comorbidity with patients with irritable bowel syndrome.

As expected, the length of treatment for the nrTMD group was significantly higher (mean = 21.9 weeks) than for the rTMD (mean = 11.6 weeks). The average length of treatment for the rTMD group was similar to what has been previously described in the literature (Dao et al., 1994). The number of treatments during therapy was also significantly higher for the nrTMD (mean = 2.3) than for the rTMD group (mean = 1.8). The results indicated that the placebo effect might have not been the only reason for treatment success as previously described (Greene & Laskin, 1982). The nrTMD group, which had more treatments, and therefore more combined placebo effects, prescribed than the rTMD, did not have any extra improvement due to that. On the other hand, the rTMD, which had less treatments prescribed, had an average pain intensity at rest reduction of 66% over baseline. The results were stable for a period of three months which is relevant.

Another supporting evidence that previous treatments and previous contacts with both medical and dental professionals may play a role in maintaining the pain, either through reinforcement of pain behaviors or learned helplessness, rather than alleviating it comes from the literature. The number of previous treatments for TMD and the number of previous clinicians consulted form the problem was a negative predictor of treatment
outcome (Fricton & Olsen, 1996; Gerschman et al., 1987; Lipton & Marbach, 1984). Gerschman et al. (1987) affirmed that: "previous behavior is the best predictor of future behavior." As previously described, our study was not specifically designed to answer these questions and our interpretations must be further supported for future research.

Pain duration has been shown in the literature to be a good predictor as well (Gale & Funch, 1984; Lipton & Marbach, 1984). However, in this study, pain duration was highly variable in both groups and had an average of 47.4 months (SD = 53.8) for the rTMD patients versus 41.6 months (SD = 45.7) for the nrTMD ones. The difference was very small and non-significant. This study disagrees with the literature on pain duration not only as a predictor but also with the actual values. The average pain duration, approximately 4 years for both TMD groups, was substantially less than the ones described by Fricton & Olsen, 1996 (mean = 8.5 years). One of the possible explanations for this difference may rely on the difficulty for the patient to remember accurately its first pain episode. One supporting evidence for this explanation is the fact that both studies had very high standard deviations showing high variability in the results.

Predictors based on long-term past pain episodes must be analyzed with caution due to the weak pain memory in chronic pain patients (Carlsson, 1983; Erskine et al., 1990).

Different from what was reported in the literature (Greene et al., 1982), the role of the treating clinician in the final treatment outcome seems also to be of less importance. The four treating clinicians had similar success rates (50 - 61.9%, total success rate of 60%); however, it must be kept in mind that all four clinicians were very well trained, experi-ent and with similar treatment philosophies. It would be interesting to compare
specialist treatment outcome versus untrained clinicians in future studies to address this issue of doctor-patient relationship properly. Comorbidity between TMD and irritable bowel syndrome (IBS) was marginally significant (p = 0.05). In addition, the nrTMD group showed a greater proportion of individuals who have been treated for IBS (27.3%) than the responding one (6.5%). This is a very strong evidence in favor of the common etiology of the chronic pain in both nrTMD and IBS patients, considering that almost one third of the rTMD also had IBS. To our knowledge, literature on comorbidity between TMD and IBS is missing; therefore, no comparisons with other studies could be made.

iv) Clinical examination variables as predictors of treatment outcome

We will now discuss the clinical variables used as predictors of treatment outcome in patients with temporomandibular disorders: a) palpation of the temporomandibular joint and masticatory muscles, b) maximum mouth opening, c) overbite and overjet, d) pain at rest and pain on chewing, e) percussion sensitivity, f) caries, and g) exacerbation of pain at rest after clinical examination.

Clinical pain measures have been shown to change from one study to the other (Goldberg et al., 1996). The results in this study on palpation of the TMJ and the masticatory muscles were very similar to the ones published by Dworkin et al. (1989, 1991) where dysfunctional (non-responding) TMD patients had higher scores on depression and somatization, but they were indistinguishable from functional TMD
(responding) patients in palpation of the masticatory muscles and the TMJ (Table V). The only predictor of was TMD treatment outcome was palpation of the posterior ligament of the TMJ in the external auditory meatus (TMJEAM). Responding TMD patients had less proportion of individuals with tenderness to palpation of the TMJ posterior ligament than non-responding ones. Despite the very high odds ratio and the statistically significant difference, it is unlikely that this finding is of any relevance. The external auditory meatus is not within the anatomical boundaries of the temporomandibular joint (TMJ)(Laskin, 1983). In addition, it has been shown that dysfunctional (non-responding) TMD patients usually present positive responses to placebo sites (Wilson et al., 1994). Therefore, this measurement also presents problems regarding reliability and validity and cannot be used as a good predictor.

The other confounder which achieved very high odds ratio (OR) was the medial pterygoid (5.5); however, the findings were non-significant. In addition to that, intra-oral palpation yielded worse reliability of measurement than the extra-oral, which makes this measurement not a good predictor. Finally, structures other than the muscle may have been palpated which compromises also our validity (Dworkin et al., 1988).

Other measurements which reached our critical OR (2.0) were the palpation of the lateral pole of the TMJ (TMJ Facial)(2.4), combined muscle score (2.0), and combined joint score (2.1). However, none reached statistical significance and the absolute difference between the two TMD groups was small. Additionally, the joint score; because it is a combined score of palpation of lateral pole of the TMJ, palpation of the posterior ligament of the TMJ as well as palpation for the detection of TMJ sounds; is also
influenced by the placebo site response (Wilson et al., 1994). Accordingly, these variables were not considered to be good predictors. All the other variables did not reach our critical odds ratio (2.0) and were non-significant.

One of the major problems in assessing TMJ and masticatory muscle palpation is the lack of good reproducibility of the results. Lobbezoo-Scholte et al. (1994) reported good percent agreement but only fair Kappa values (Landis & Koch, 1977): total joint noises (80%, 0.23), masseter muscle (90%, NA), temporal muscle (71%, 0.36), medial pterygoid muscle (67%, 0.35), TMJ (65%, 0.30). Kopp (1977) also found similar results for masticatory muscle palpation (Scott’s pi and percent agreement): medial pterygoid muscle (0.42, 78%), superficial masseter muscle (0.33, 56%), deep masseter muscle (0.32, 81%), lateral pterygoid muscle (0.25, 47%), anterior temporal muscle (0.22, 46%), insertion of the temporal muscle (0.13, 46%), TMJ laterally (0.32, 77%), TMJ external auditory meatus (0.15, 77%). Carlsson et al. (1980) also showed similar results (percent agreement and Scott’s pi): palpation of the TMJ (70% and 0.36) and masticatory muscles (80% and 0.69).

Some studies have been able to reduce the variability in the results by calibrating examiners. DeWijer et al. (1995) found the interexaminer reliability of the tests measuring maximal active mouth opening (k=0.56) and registration of clicking during active mouth opening was high (k=0.70). The interexaminer reliability was fair for the tests measuring the intensity of pain during active movements and moderate for tests recording joint sounds (k = 0.47 to 0.59). Palpation of masticatory muscles was fair (k=0.16 to 0.45). There was high interobserver agreement on several items of the traction
and translation tests, although the kappa values were low. The interexaminer reliability of the multitest scores for compression was substantial for joint sounds (k = 0.66) and fair for pain (k = 0.40). The interexaminer reliability of the multitest scores for muscle palpation and joint palpation was moderate (k = 0.51) and fair (k = 0.33), respectively.

Another problem which makes difficult to assess predictability of TMJ and masticatory muscle palpation is the fluctuation of signs and symptoms over time. Kopp (1977) showed that the fluctuation of clinical signs during a six-week period was considerable in patients with TMD who received no other treatment other than counseling. Conversely, in the study by Carlsson et al. (1980), the clinical signs were much more constant during a period of five weeks in an epidemiological sample of children without severe signs and symptoms or dysfunction. According to the authors, variation in clinical signs can - besides from actual changes - be due to poor precision or reproducibility in the methods used, or, if more than one observer is involved, to differences in examination technique and different opinions about positive and negative findings.

The overall results in this investigation, similar to the ones by Dworkin et al. (1989, 1991), seem not to support palpation of the temporomandibular joint and masticatory muscles as good predictors. However, in most cases, nrTMD patients were more sensitive to palpation than rTMD ones. These findings for TMJ and masticatory muscles were similar to the ones described by of Goldberg et al. (1996) for the idiopathic TMD group. They also found that the post-traumatic TMD group had higher proportion of patients sensitive to palpation on the masticatory muscles and TMJ than the idiopathic
TMD one. However, in the latter study this difference was found to be significant for the masseter, temporalis and sternocleidomastoid which was not observed in this investigation between the two TMD groups. Additionally, some relevant odds ratio were found for some variables. Notwithstanding, the significance and relevance of this finding is difficult to assess considering that the sample size was not calculated and our study not intended to answer this question.

A single examiner was used, because the clinical examination was only done once at baseline and not at follow-up which does not require intra-examinater calibration. Furthermore, the fluctuation of signs and symptoms after six months would have made ineffective our calibration procedures. At least, most studies demonstrated a higher intra-observer than inter-observer consistency, and the conclusion was that longitudinal studies or treatment evaluations should be performed by the same observer (Carlsson et al, 1980; Kopp & Wenneberg, 1983; Fricton & Schiffman, 1986).

Maximum mouth opening, overbite and overjet did not show any statistically significant difference between the two TMD groups and was not considered a good predictor (Table VI). Their absolute differences were also very small and not conclusive. The results in this study were also similar to the ones by Goldberg et al. (1996) who found that none of the patients in the idiopathic TMD group had interincisal opening less than 35 mm and could not be distinguished from the post-traumatic TMD based on that. In this investigation, only 5% of the overall TMD population had less than 35 mm. Our study was also in agreement with Dworkin (1989, 1991) who found that both functional and dysfunctional TMD patients could not be distinguished by unassisted vertical range
jaw motion. It must be kept in mind that the reliability of measurements of mandibular range of motion, similarly to TMJ and masticatory muscle palpation, is moderate. Kopp (1977) reported the Scott's pi and percent agreement for mandibular range of motion were the following: a) maximal mouth opening (0.42, 72%), b) protrusion (0.40, 88%), laterotrusion to the right (0.32, 87%), and pain on laterotrusion to the left (0.19, 79%). Carlsson et al. (1980), also found similar results for mandibular range of motion: 83% and 0.44 and DeWijer et al. (1995) found also moderate Kappa results (Landis & Koch, 1977) for maximum mouth opening (0.56). Therefore, based on the results of this study and the literature, mandibular range of motion seem not to be good predictors of TMD treatment outcome.

Pain at rest pre-treatment was also similar between the two TMD groups (range 63-64 mm on a 100 mm VAS scale) and comparable to previous studies (56 mm) (Linton & Melin, 1982); however, there was a significant reduction in pain at rest post-treatment in the rTMD patients when compared to baseline (66%) and was statistically significant different than the nrTMD ones which had only 4.7% improvement over baseline. Comparison with previous studies were made in the pain improvement section of the discussion.

One interesting finding was that pain on chewing pre-treatment was statistically significantly higher (60 mm on 100 mm VAS) in the nrTMD than in rTMD patients (39 mm). Analogously, the average reduction in pain on chewing over baseline scores (36%) was statistically significantly higher in the rTMD group as compared to the nrTMD (0%). This aggravation of pain during functioning has also been reported by Dao et al. (1994),
who found that 67.2% of the patients reported pain during mastication. In addition, pain intensity and pain unpleasantness increased 55.6% and 62.3%, respectively; when the same population was asked to chew on half a leaf of green casting was gauge 28 (Kerr) for 3 minutes. Dworkin et al. (1989) classified non-responding TMD patients as dysfunctional due to their inability to cope with TMD pain, and Schnurr et al. (1991) found also ability to cope as a good predictor of TMD treatment outcome. This study also agrees with those reports, because similar levels of pain at rest at baseline between the two TMD groups had different impacts on levels of functioning also at baseline. It must be pointed out that this study did not evaluate the overall mandibular functioning and quality of life, and the interpretations of our findings are limited. Our study further corroborates the findings of Freeman et al. (1998), where the presence of a psychopathologic disorder correlated with significant differences in the patient’s perception of their pain and limitation in jaw opening. This is crucial when we try to understand why non-responding patients have greater perceived disability and psychosocial impact in their normal functions, as in perceived pain on chewing, than the responding ones. In the study by Freeman et al (1998), the TMD patients were evaluated after arthroscopic surgery for the pain and limitation of mouth opening. They also demonstrated that pain was a more relevant outcome measure than maximal mandibular opening after surgery, which remained almost unchanged after treatment for both responding and non-responding groups. Further, in this investigation the authors found that, at least initially, patients perceived improvements in maximal opening while this could not be shown objectively. Interestingly, the patients with psychopathological
disorders ultimately began to perceive not only increased pain but also worsening of mandibular opening after about three months while this did not happen in patients who did not have obvious psychopathological disorders. This suggested that, as in our study, the sub-group of patients with psychopathological disorders continued to recognize or be aware of pain and dysfunction that was actually still present following treatment while patients who did not have psychopathological disorders could effectively “ignore” their pain and dysfunction following treatment.

The other variables included in the intra-oral examination (i.e. percussion sensitivity, caries, exacerbation of pain after examination) were non-significant and did not reach our critical odds ratio (2.0). With the exception of pain exacerbation after examination, most patients had negative findings. Pain exacerbation was positive for both rTMD and nrTMD groups (88 and 91%, respectively) and justified our choice of performing the neuropsychological and psychosocial tests prior to clinical examination; otherwise, this increase in pain levels would have most likely affected our test results.

v) Assessing neuropsychological predictors controlling for confounders

In the logistic regression analysis, all neuropsychological tests used, even the ones which were negative predictors were included, considering that a confounder may act not only creating false positives, but also false negatives (Norusis, 1991, 1992). On the total, eleven confounders which were not controlled in the design stage (e.g. gender, language,
pain diurnal variation and place of testing) were included in our logistic regression analysis (e.g., age, fatigue, energy level, Sleep Assessment Questionnaire, Beck Depression Inventory, educational level, income, length of treatment, pain duration, pain at rest pre-treatment and pain on chewing pre-treatment). The choice of these confounders has already been outlined in the results section. The confounders were analyzed one at a time to assess the influence of each variable in the initial odds ratio (y) and also all confounders at the same time.

The range of percent correct observations was high (76.1 to 85.7%) and comparable to other studies. Fricton & Olsen (1996) created a regression model which was capable of predicting 87% of the successful cases. Lipton & Marbach, 1984, reported that the eight factors included in their discriminant function allowed them to correctly classify 79% of temporomandibular subjects. Gale & Funch (1984) were able to predict 83% of the short-term successes and 100% of the long-term ones. Finally, Millstein-Prentky & Olson (1979), using discriminant scale, was able to correctly predict the treatment outcome of 93% of the successful patients and 88% of the unsuccessful ones.

From Table VII, it could be determined that none of the eleven confounders included in our logistic regression analysis influenced none of the neuropsychological tests in any direction (false positives or false negatives). Consequently, the neuropsychological tests results described on Table III were unconfounded for all variables included in the analysis. The variable 'types of treatment' could not be included as discussed in the results section, but the other related variables (e.g. success rate among different clinicians as well as number and length of treatments) were either non-
significant or did not change the outcome in the logistic regression analysis. Therefore, it is unlikely that the different treatments provided (i.e. full-coverage lower splint, cyclobenzaprine, diflunisal or physiotherapy) would have had any impact on the outcome. In conclusion, the CVLTCR, CVLTCL and the CCC are still good predictors for TMD treatment outcome after controlling for confounders.

On Table VIII, the some of the neuropsychological and psychosocial tests which have shown to be good predictors of TMD treatment outcome on Table III were recoded to allow the calculation of the odds ratio. The CVLTCR, the CCC, the SAQ and the BDI were dichotomized. The rationale for the choice of the tests and the cutoff points was explained in the results section.

Regarding the CVLTCR, it can be determined that patients who have scores lower or equal to 44 in the CVLTCR will have 3.4 times (critical OR = 2.0) the probability of becoming nrTMD patients. One of the advantages of using the CVLT as a predictor is the fact that the test has published normative data (age- and sex-matched) (Delis et al., 1987), which makes it easier to compared our patient values to external controls. The CCC had an even higher probability (OR = 4.5) for patients who had scores equal or below 30 to become a nrTMD patient. The CCC, similarly to the CVLT, also has published norms for different age and gender groups (Stuss et al., 1985, 1987, 1988). One of the advantages of the CCC over the CVLT is the fact that the test seems to vary less among patients across different age and gender groups. In addition, it has also demonstrated to have higher odds ratio than the CVLTCR, and therefore, it seems to be a better predictor. The final
assessment will be discussed later in the discussion of our stepwise logistic regression procedure (Table XVI).

Positive scores (i.e. greater or equal 16) in the SAQ increased 5 times the probability of TMD patients to become non-responders (cutoff = 16), which was the highest among all predictors. Sleep habits have already been described in the literature as a good predictor (Fricton & Olsen, 1996), but to our knowledge, this is the first study which uses a reliable and valid sleep questionnaire. The dental literature has been overlooking a more close association between sleep disorders and TMD, which has been better studied in the medical literature for other chronic pain conditions such as fibromyalgia and chronic fatigue syndrome (Côte & Moldofsky, 1987, Moldofsky, 1993a, 1993b). This study used sleep assessment as a confounder for neuropsychological tests, and further studies must be designed specifically for assessment of sleep as a perpetuating factor in TMD patients.

The BDI, despite being marginally significant on Table III (p = 0.08), increased 3.1 times the probability of TMD patients of becoming non-responders, if the patient had a score greater or equal to 16. The wide variation in the odds ratio may be an indication that our sample size was not large enough to detect significant differences in the BDI. The results in this investigation also agree with previous report which found depression to be a good predictor (Dworkin et al., 1989, 1991; Nielsen & Williams, 1980; Schwartz et al., 1979). Considering that depression has been correlated with other psychosocial variables (e.g. low self-esteem, low energy, feeling worried, and sleep)(Fricton & Olsen, 1996),
unconfounded studies on the role of depression in the perpetuation of TMD signs and symptoms must still be carried out.

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were not ideal (i.e. close to 0.75) for any of the tests (Dworkin & LeResche, 1992). From Table VIII, we found that the CVLTCR had higher specificity and better ability to detect rTMD than the CCC. On the other hand, the CCC had better sensitivity and ability to detect nTMD than the CVLTCR. Both tests had moderate NPV and PPV (0.63 - 0.72) and were reasonably good to predict both rTMD and nTMD patients once the test is diagnosed as positive or negative. The SAQ had excellent capacity to differentiate nTMD; but very low for rTMD. Conversely, the ability to predict disease after a positive test result was very low, but very high for non-disease prediction. Finally, the BDI, similar to the SAQ, had high ability to detect nTMD patients, but low for rTMD ones. On the other hand, the ability to predict disease and non-disease after a positive or negative test result was moderate. It must be pointed out that none of the tests can be used in isolation, and the test results must always be analyzed within the context of other factors collected during the clinical examination. Considering that our major objective is to detect nTMD patients, the SAQ seems to be more useful than the other three tests due to its very high specificity (0.91). However, these findings must be confirmed in specifically designed studies for sleep and depression.

In our study, among all neuropsychological test predictors, only the CCC in the majority of the nTMD population was actually outside of the normal range (58.3%), which would genuinely characterize a neuropsychological ‘deficit’ in the true meaning of
the word for that test. Therefore, despite having significantly different neuropsychological test scores when compared to other groups, both the nrTMD and the IBS groups are not outside of the normal range for both chronic pain and non-pain populations. Among the psychosocial tests, the SAQ yielded the most surprising scores: 78% of the overall TMD, 68.6% of the nrTMD, 91.7% of the nrTMD, and 80% of the IBS population were outside of the normal range. In addition, 27% of the non-pain population was also scored outside of the normal range. These findings suggest strongly that sleep disorders are more frequent than initially though, not only in the general population, but also in chronic pain populations. In addition, the SAQ was the only test capable of differentiating the chronic pain versus the non-pain populations.

In synthesis, the California Verbal Leaning Tests (CVLTCR and CVLTCL) and the Brown-Peterson Consonant Trigram Auditory Memory Task (CCC) have shown to be good predictors of treatment outcome in patients with temporomandibular disorders (TMD). This findings remained stable even after controlling for thirteen confounders (i.e. age, fatigue, energy level, sleep, depression, educational level, income, number of treatments provided, length of treatment, comorbidity with irritable bowel syndrome, pain duration, pain at rest pre-treatment, pain on chewing pre-treatment). Some psychosocial variables, despite not being controlled for confounders and not being our research question, which have shown to be good predictors were sleep, fatigue and income. Other uncontrolled variables, such as pain on chewing and at rest pre-treatment were also found to be good predictors. The other clinical variables usually assessed during TMD examination (e.g. mandibular range of motion, TMJ and masticatory muscle palpation,
overjet, overbite, percussion sensitivity, and pain exacerbation after examination) were not considered good predictors. The sociodemographic description of both TMD groups did not differ significantly from what has been reported in the literature.

vi) **Selection of the best predictors among the predictors of TMD treatment outcome**

In this investigation, stepwise (backward elimination) logistic regression was used to identify within a subset of our predictors of TMD treatment outcome, which were the ones capable of being the best predictors (Figure 2). Similar techniques have been described in the literature (Norusis, 1991, 1992; Fricton & Olsen, 1996). The model used the neuropsychological, psychosocial, and confounding variables which proved to be the best predictors for TMD (i.e. CVLTCR, CVLTCL, CCC, fatigue, energy level, depression, sleep, educational level, income, pain duration, pain at rest pre-treatment, and pain on chewing pre-treatment). In addition, length of treatment, despite being assessed during the course of treatment, was also included due to its statistically significant difference between the two TMD groups (Table IV).

In the backward elimination procedure; where all variables were included in the model, and the ones which were not good predictors were eliminated one at a time; was used. The best predictors were the CCC, fatigue and length of treatment. The model was capable of predicting successfully (percent agreement = 82%) the majority of the
temporomandibular treatment outcomes. The sensitivity and specificity were also comparatively high (0.78 and 0.85, respectively) and higher than the reported critical value for both (0.75) (Hennekens & Buring, 1987). The percent agreement in this investigation was also comparable to one by Fricton & Olsen (1996) of 80%.

Fatigue and length of treatment were also good predictors in the bivariate analyses; nevertheless, length of treatment could only be assessed during treatment itself and cannot be used, and normative data for fatigue are missing. In addition, fatigue was found to be closely associated with depression and sleep problems (Appendix IV). Besides, normative data for fatigue has not been published which makes difficult to used it as a diagnostic test. Therefore, their use as predictors must be reassessed in a study specifically designed to answer these questions.

Among the neuropsychological tests, the Brown-Peterson Consonant Trigram Auditory Memory Task (CCC) proved to be the best unconfounded predictor in both bivariate and multivariate (stepwise) analyses. In addition, the test has already published norms and proved to be not very affected by age or gender (Stuss et al., 1985, 1987).

e) Associations between the responding and non-responding TMD vs. the IBS groups

The findings related to our second hypothesis will now be discussed. We will first start discussing the social and demographic distribution of IBS in comparison with previous studies and our own. Then, we will discuss the findings between the nrTMD
group versus the IBS, then we will proceed with the findings between the rTMD group versus IBS. Clinical variables (e.g. pain intensity and pain duration) were not analyzed, because IBS patients did not undergo clinical examination. After that, the clinical implications of those findings will also be analyzed.

i) Social and demographic distribution variables

The four confounders which were not controlled in the design stage (i.e. educational level, employment, income, age) showed no statistically significant difference between nrTMD versus IBS patients (Table X). Similarly, no statistically significant difference was found between rTMD versus IBS patients (Table XIII). Notwithstanding, educational level was slightly higher for the IBS and rTMD groups. Employment was a little higher for the rTMD group than the other two. IBS had higher levels of income and average age than the TMD groups. Therefore, it is unlikely that these variables may have played a role in our neuropsychological tests; however, this will be properly addressed further in the discussion of the logistic regression analysis. The socioeconomic and demographic distribution for IBS is similar to other studies in the medical literature (Drossman et al., 1994, 1995). These studies reported that IBS patients tended to be between the average ages of 39 to 41, predominantly women (92%), well educated (average 12 to 15 years of education), with middle to high income level (US$ 35,000 or above) and employed (59%). The only difference was that our average age (32.9 years)
was a little lower than previously described; however, this may have been only a function of our limited sample size (n = 20).

ii) Neuropsychological and psychosocial tests

In all reaction time tests (SRT and MCRTs), no statistically significant difference was found between nrTMD and IBS groups. IBS had lower reaction time test results than the nrTMD patients, but the actual differences observed between the two groups in all reaction time tests was smaller than 100 msec, which was not considered to be relevant by Stuss et al. (1985, 1989a, 1989b)(Table IX). No statistically significant difference was also found in any of the tests which evaluates attention and short-term and short-term memory under interference (CVLTs and CCC), and the actual difference in all tests was also extremely small. Similarly, the psychosocial tests (fatigue and energy level, BDI and the SAQ) showed also similar and non-significant differences between the two groups. Nevertheless, in this study, the IBS group had slightly higher levels of sleep disorders, depression, fatigue and lower energy level. However, the proportion of patients with sleep disorders (cutoff = 16) was higher for the nrTMD group. Sleep disorders were found in 80% of the IBS population and 91.7% of the nrTMD groups; and depression (cutoff = 15) in 45% and 29.2%, respectively. This may raise the hypothesis whether the IBS, despite having similar pain levels, has or not greater impact on pain functioning than the nrTMD group (Von Korff et al., 1988). However, our study was not designed to measure levels of
disability between and among the four different groups and must be tested in future studies.

Table XII showed the results for the neuropsychological and psychosocial tests between the rTMD and IBS groups. The results were extremely similar to Table III (rTMD versus nrTMD), where statistically significant difference between the two groups were found for the following tests: a) CVLTCR, b) CVLTCL, c) CCC, d) SAQ and e) fatigue. In addition to that, on Table XII, two extra tests were also found significant: a) BDI and b) energy level. Similar to what was described for the two TMD groups, IBS had worse scores in the neuropsychological tests (with the exception of the reaction time tests and CVLTP), higher levels of fatigue, depression and sleep disorders, and lower energy level. Sleep disorders (cutoff = 16) were found in 80% of the IBS population versus 68.6% of the rTMD group; and depression (cutoff = 15) in 45% and 11.8%, respectively.

In the logistic regression analysis for nrTMD versus IBS and rTMD versus IBS groups (Tables XI and XIV, respectively), seven confounders were included (e.g. age, fatigue, energy level, SAQ, BDI, educational level and income). This was done in order to assess if our neuropsychological tests results from Tables IX and XII were still valid after controlling for confounders. Changes in the level of odds ratio for all neuropsychological tests were done for all tests to detect any difference greater than 15%. For almost all neuropsychological tests, the results remained unchanged after controlling for one confounder at a time and also all at the same time. Changes were only detected for the CVLT on Table XI, which became a positive test under the influence of all confounders; and on Table XIV, where the individual influence of SAQ and BDI also changed it into a
positive test result. It must be pointed out that the low scores on the CVLTI (between 0 and 1) may have been responsible for this finding due to an increase in data variability; and therefore, it was not considered to be meaningful. Therefore, the positive test results for both Tables IX and XII remained stable controlling for all confounders.

The results in this investigation confirmed the similarities which have been reported between TMD and irritable bowel syndrome. Irritable bowel syndrome, similar to TMD, is more prevalent among women, have decreased prevalence with age, and is probably self-limiting (Gerke et al, 1988). It also confirmed previous reports that irritable bowel syndrome and TMD have been associated with affective disorders (i.e. myalgia, fatigue, unrefreshing sleep and emotional distress) which were also common with other chronic pain conditions, such as chronic fatigue and myofascial pain syndrome (Hudson et al, 1992; Moldofsky, 1993a).

The findings in this study were very similar to ones described by Von Korff et al. (1988). In an epidemiological comparison of pain complains, they found that headaches, abdominal pain (IBS), chest pain and facial pain (TMD) had similar pain prevalence in the prior six months, similar pain intensity levels, and number of average lost days in the last 6 months. All pain conditions were typically long standing, recurrent of mild to moderate intensity and usually did not limit activities. They also found that chronic pain patients had higher levels of anxiety, depression, and non-pain somatic symptoms as measured by the Symptom Checklist (SCL); poorer self-rating of health status; and more family stress compared to persons without a pain condition. The only difference was the fact that in this investigation, the IBS group had slightly higher sleep disorders,
depression, and energy levels than the nrTMD patients; however, they were all non-
significant. Finally, these chronic pain conditions were associated with similar impact on
psychosocial functioning and health care costs (Dworkin, 1994).

In summary, the neuropsychological and psychosocial profile of both IBS and
nrTMD patients were indistinguishable from one another and similarly different than the
rTMD group, which indicated that these two chronic pain conditions were similar to each
other as far as neuropsychological and psychosocial assessment is concerned. The
irritable bowel syndrome and nrTMD groups differed in exactly the same
neuropsychological and psychosocial tests as with the rTMD patients (CVLTCR,
CVLTCL, CCC, sleep and fatigue). They differed however in depression and energy
levels, where IBS patients also differed significantly from rTMD. Despite the similarity
between the nrTMD and the IBS groups, the levels of depression and sleep were higher
while energy level was lower, despite non-significant, in the IBS group. The description
of the IBS population did not differ significantly to what has been reported in the
literature.

f) The non-pain group (Group IV) versus TMD and IBS patients

The results on Table XV were not assessed for confounders, because this was not
our hypothesis and were used only to generate hypotheses. Our most important findings
came from the differences between the four different groups rather than among them,
where they simply show the same differences already explained on previous Tables (III, IX, XII). Therefore, we will concentrate our discussion in the comparisons between the non-pain versus the patient populations.

In this investigation, language-, age- and sex-matched non-pain comparison group (internal control) showed no statistically significant difference in the social and demographic variables with the other patient groups (Table I) which made our neuropsychological and psychosocial results comparable, despite not controlled.

In the reaction time tests, the non-pain population values were comparable to published normative data (Stuss et al., 1989a, 1989b). All test results were comparable, but the non-pain group tended to be a little faster than the other three groups, but the differences ranged between (1 to 80 msec) which at times is higher than the sample size estimate (75 msec) but considered to be non-relevant (Stuss et al., 1985, 1989a, 1989b). This helps to clarify one question raised by Goldberg et al. (1996) regarding how the two chronic populations studied (post-traumatic and idiopathic TMD) would compare to an “asymptomatic control group.” He did not have internal controls and only compared his data to external ones. This study agrees with their statement that the reaction time tests in the post-traumatic TMD was substantially lower than asymptomatic groups; however, this study could not substantiate his observation that the idiopathic TMD group also had apparently lower reaction time level. In this investigation, the neuropsychological results were slightly lower, but neither significant nor relevant when compared to the patient population. Therefore, our findings could not substantiate reaction time test not only as a
good predictor of treatment outcome for idiopathic TMD patients, but also as a good diagnostic test to differentiate patient versus non-patient populations.

Similar to the reaction time tests, the non-pain values for the remaining neuropsychological tests (i.e. CVLTCR, CCC, SAQ and BDI) were similar to normative data from previous studies (Beck, 1970; Delis et al., 1987; Stuss et al., 1985, 1987; Cesta, Moldofsky & Sammut, 1996). Unfortunately, normative data for the CVLTCL, CVLTP, CVLTI, fatigue and energy level were not available. This investigation found statistically significant differences between the non-pain group and the IBS patients in some of the neuropsychological and psychosocial tests (i.e. CVLTCR, CCC, fatigue, energy level, SAQ, and BDI). Analogously to IBS, the non-pain group showed exactly the same differences in the neuropsychological and psychosocial test results with the nrTMD group. The only exception was found in the energy level, where the non-pain population had higher values than both the nrTMD and the IBS groups; however, the difference was only significant for the IBS population.

In all but one neuropsychological and psychosocial tests, the non-pain population showed no statistically significant differences with the rTMD group. However, the CVLTCR, CVLTP, CCC, fatigue, energy level, and depression scores, despite non-significant, were worse for the nrTMD group as compared to the non-pain population. The SAQ was the only statistically significant test between the two groups. This also stressed the importance of further studies on sleep disorders, because this was the only test not only capable of differentiating our three chronic pain populations, with the
exception of IBS versus nrTMD, but also of differentiating our non-pain from our chronic pain populations.

The findings in this investigation strongly suggest that the neuropsychological scores from a non-pain population are more similar to the rTMD group, despite slightly higher, than to both the nrTMD and IBS groups. Notwithstanding, our sample size was not calculated to answer these questions and must be analyzed with caution. Finally, it should also be stressed that these findings were not controlled for confounders and must be used only as hypotheses generating.

g) Sources of biases, clinical implications and future considerations

Systematic errors and limitations are always present in any study evaluating particularly chronic pain populations. The predictors selected for our TMD and IBS populations cannot be extrapolated to other chronic pain populations, despite the similarities found between the nrTMD and IBS groups. In addition, our limited sample size and inclusive criteria limits our generalizability. Further studies should include more than two chronic pain populations and more than two TMD sub-groups.

In pain assessment, cultural differences and the context of the situation may have also played a role (Chapman et al., 1985; Gerschman et al., 1987). In this study, despite assessing pain in a multicultural setting, the subjects were also matched by language. Therefore, we tried to minimize the effect of cultural differences by using simple to
understand unidimensional pain assessment (100 mm VAS, pain at rest) and patient’s self-perception (better, worse or same) as well as by including only patients who demonstrated fluency in English. Questions have been asked in the study regarding its objectives, but all patients demonstrated good understanding about its instructions. In no case, the test had to be interrupted due to difficulty in either understanding or responding a specific task. Finally, we only assessed the TMD patients at six-month follow-up, but the literature suggests that TMD management also has good long-term success rate (Greene & Laskin, 1983).

The improvement rate (60%) in this study is similar to the one reported by Fricton & Olsen (1996), but it is lower than the average reported in the literature (70-98%, Greene & Laskin, 1983), indicating that our inclusion/exclusion criteria (Dworkin & LeResche, 1992; Dao et al., 1994) was very restrictive. However, this very restrictive criteria helped us to better defined TMD groups and sub-groups and avoid selection biases.

In addition, the proportion of nrTMD cases was higher (40% or approximately 2:3) than previously reported (1:3 to 1:9, McNeill, 1985; Centore et al., 1989). However, it was unlikely that this may have affected the test results, because these two TMD sub-groups were assessed separately. Notwithstanding, the results may have affected positively our reported rates of depression and sleep for the TMD population as a whole, considering that the scores for the nrTMD population were higher than the rTMD group. Nonetheless, it did not affect the sub-groups and related neuropsychological tests analyzed. In addition, the sociodemographic variables described did not show any
difference between the combined TMD group and what has been reported in the literature (Fricton & Olsen, 1996; Schnurr et al., 1991; Gerschman et al., 1987). Therefore, the higher proportion of nrTMD patients also did not affect our sociodemographic composition.

The neuropsychological tests were either computer driven, or they were read to the patient and scored in a highly standardized fashion, including time of the test and test setting. Therefore, it is very unlikely that patient and examiner biases may have played any substantial role. The psychosocial variables can also be affected by different cultural interpretations, but our results for depression and sleep were comparable to normative standards and TMD studies for the non-pain and TMD groups (Beck, 1970; Cesta, Moldofsky & Sammut, 1996; Gerschman et al., 1987).

In clinical examination with palpation of specific muscle bands or ligaments, as reported by Fricton & Olsen (1996), error is introduced when there is variability in the amount of pressure applied, the palpation technique, the size of the distal phalanx, and the area actually palpated. Additionally, our TMD clinical examination relied on a trained but not calibrated single examiner which may have introduced examiner biases. However, our procedures were blinded to outcome, and the results were comparable to previous ones (Dworkin et al., 1989, 1991). However, the lack of agreement in some cases between this investigation and the one by Goldberg et al. (1996), who used a similar clinical examination form, may have been determined by the lack of calibration between the two examiners. Nevertheless, our overall results were also similar.
One of the limitations of our study is the fact that it is still not clear to what extent IBS and TMD, and most likely other pain conditions, symptoms represent normal perception of abnormal function or abnormal perception of normal function. It must be emphasized that the presence of cognitive impairment between the two TMD sub-groups themselves and also with the IBS and non-pain populations does not establish *per se* a cause and effect relationship. However, what the study suggests is that neuropsychological tests can be used not only as predictors of TMD treatment outcome but also a solid evidence of comorbidity between the chronic pain conditions studied. Indeed, the two TMD sub-groups had different proportions of comorbidity with IBS, with nrTMD patients having substantially higher proportion of patients who have been treated for IBS than the rTMD group.

Other limitations were that neuropsychological and psychosocial testing as well as clinical examination were not performed at follow-up due to logistic reasons and budget limitations, only assessment of pain at rest and pain on chewing. Also, only few neuropsychological tests were used due to the same reasons described above. In addition, neuropsychological tests must also be re-assessed at follow-up in order to determine if, besides pain improvement, TMD patients and other chronic pain conditions also improve in terms of neuropsychological performance.

Another limitation was that the limited number of neuropsychological tests were used. Further studies using more complex batteries of neuropsychological tests (Spreen & Strauss, 1991) are still needed in order to confirm the present findings and expand our
knowledge of the breadth and depth of cognitive impairment in TMD as well as other chronic pain conditions.

There are major clinical implications regarding the findings of this study. Initially, it must be stressed again that neuropsychological differences are not proof of etiology. Our results simply indicate that neuropsychological tests can be used as predictors of treatment outcome in non-traumatic TMD patients and this does not necessarily imply a cause and effect relationship. Therefore, at face value, the neuropsychological tests must be considered risk indicators, rather than risk factors for TMD. In addition, neuropsychological performance was closely correlated with psychosocial scores; i.e. patients who performed poorly in the neuropsychological tests tended to have lower levels of energy, and higher levels of depression, fatigue and sleep. However, once again, a cause and effect relationship was not established.

The nrTMD and IBS groups had similar neuropsychological and psychosocial test results, performing poorly when compared to the rTMD and non-pain population. One interesting fact was that the reaction time tests, which were significantly higher for a post-traumatic TMD population in a previous study (Goldberg et al., 1996) when compared to the non-traumatic TMD group, were not capable of distinguishing any of the four groups here. What is even more intriguing was the fact that post-traumatic TMD neuropsychological scores were comparable to patients with closed head injury (Stuss et al. 1985, 1989a, 1989b).

Taken together, these data suggest that, based on the neuropsychological test performance, post-traumatic TMD patients, and those with closed head injury might have
similar underlying etiologic factors perhaps pointing to some degree of mild concussion. This would explain why a basic and simple task like the reaction time tests, which demand only attention and velocity to react to a stimulus, were affected in these two groups in the previous study, but not in this one. On the other hand, our data suggest that affective disorders, which were higher for the nrTMD and IBS groups when compared to the rTMD and non-pain populations, may have influenced the neuropsychological tests which assess predominantly memory which is a higher and more complex function demanding high levels of attention and information processing (i.e. verbal memory, short-term memory, and short-term memory under interference) (Stuss et al., 1985, 1989a, 1989b).

Notably, there may be corroborative evidence for these suppositions found in other investigations pointing to neurophysiological mechanisms for non-responsive pain. Davis et al. (1997), using functional MRI, reported that the pain and attention areas in the anterior cingulate cortex (ACC) were closely localized. The pain activation areas in the ACC were in the anterior part; while, the areas involved with attention-demanding cognitive tasks were located in the posterior part of the ACC. What is even more noteworthy is the fact that the ACC, similar to the medial thalamus and limbic areas, is generally considered part of the motivational-affective pain system, mainly because it receives input from medial thalamic nuclei (Sikes and Vogt, 1992). Moreover, it has been shown that surgical cingulotomy may reduce chronic intractable cancer pain, indicating the importance of this area for the processing of information regarding perception of chronic pain (Bouckoms, 1989; Pillay and Hassenbusch, 1992). This neural
pathway may thus provide an explanation as to why higher levels of depression, sleep and fatigue as well as lower levels of energy can result in lower performance in more complex cognitive tasks.

The ACC also receives input from the spinothalamocortical pathway that terminates in somatosensory cortex, which has been associated with the sensory-discriminative aspects of pain (Albe-Fessard et al., 1985; Apkarian, 1995). This pathway may also help us to understand why the nTMD patients also perceive more muscular and TMJ pain as well as pain on chewing when compared to the rTMD group. Given these issues, it is possible that these inputs from the muscles and joints seem to have less impact in the neuropsychological performance than the affective component, but cannot be “ignored” as suggested above in relation to the study reported by Freeman et al (1998). This may also explain why the classical signs of TMD such as muscle tenderness would not be and in fact could not be predictors of treatment outcome. Our findings, in light of the fMRI, data may thus suggest that recovery depends not on elimination of muscle pain or limited opening but rather the ability of a cognitively intact individual to, in effect, “learn to ignore” the painful muscles and mandibular limitation (perhaps by bypassing or otherwise not activating the attention center). Individuals who are less cognitively “adap" may not be able to perform this learned skill and are thus doomed to be aware of their peripheral pains, wherever they may be unless more creative treatment approaches are investigated as will be discussed below.

These findings have major implications in the management of chronic pain conditions. Considering that chronic pain conditions are multifactorial in origin,
multidisciplinary teams addressing a variety of conditions probably is the most appropriate way to these cases. The literature has shown that cognitive-behavioral treatment (CBT) helps patients with IBS to increase recognition of the role played by attention allocation, personal appraisal style and illness beliefs in chronic pain and psychosomatic disorders. CBT which is useful for IBS management, may also be possibly helpful with other chronic pain disorders (Toner, 1994). However, randomized controlled clinical trials (RCT) of CBT for chronic pain conditions are still missing. This is particular true for IBS and non-responding TMD, because IBS is a predominantly a non-responding condition. Other forms of affective disorders management may also be employed; relaxation/stress therapy has been shown also effective in both IBS and TMD patients, because of its effect in reducing autonomic arousal and anxiety (Turk et al., 1993b; Toner, 1994). Analogously to CBT, there is a lack of RCTs to confirm the treatment efficacy of both managements. Finally, considering that sleep was our best predictor among the psychosocial tests, randomized controlled trials using particular sleep medications (e.g. Imovane, Zopiclone) must be carried out not only in nRTMD but also in IBS patients.

As far as clinical practice is concerned, TMD patients with history of motor vehicle accidents must be tested for reaction time tests. If the test results are not considered to be significantly different than external controls, than the referral to a neurologist is recommended. If not, than the patient can be treated by a dentist or specialist. In addition, memory tests, in particular the CCC, should be indicated for TMD patients in order to disclose the cases with good versus poor prognosis. Those with scores
which are significantly different from controls could be referred for cognitive behavioral therapy, rather than trying reversible treatments that have been shown to be ineffective in these patients and may even cause an increase in the pain level (Fricton & Olsen, 1996). The literature has shown that cognitive behavioral therapy is effective in IBS patients; however, randomized controlled trials must also be done in order to assess the treatment efficacy for nrTMD patients.
VIII. CONCLUSIONS

In this investigation, clinical and psychosocial variables were tested as predictors of treatment outcome for patients with temporomandibular disorders (TMD). In addition, this investigation makes a unique contribution which was missing in the dental literature, we also included neuropsychological tests as predictors of treatment outcome in TMD patients without history of motor vehicle accidents.

Based on some of the neuropsychological test results; the California Verbal Learning Test - correct responses (CVLT CR), the California Verbal Learning Test - clusters (CVLTCL), and Brown-Peterson Consonant Trigram Test (CCC); we were capable of differentiating responding (rTMD) versus non-responding TMD (nrTMD) patients prior to treatment. Our results remained valid even after controlling for thirteen psychosocial clinical variables used in the analysis and other four included in the designed (e.g. gender, language, time of testing, site of testing, age, fatigue, energy level, sleep, depression, educational level, income, number of treatments provided, length of treatment, comorbidity with IBS, pain duration, pain at rest pre-treatment, and pain on chewing pre-treatment). Reaction time tests were not found to be good predictors. Among the best predictors, the CCC in a backward stepwise logistic regression was chosen as the best predictor of all neuropsychological tests.
Three psychosocial variables were also found to be good predictors: a) sleep, b) depression, c) fatigue, and d) income. Sleep was particularly useful, because it was capable not only to differentiate different chronic pain conditions, but also to differentiate non-pain from chronic pain populations. It must be accentuated that despite the fact that on average the groups which had worse neuropsychological performance also had worse scores of sleep, depression, fatigue and energy level; this in itself is not evidence of causation, and this relationships must be further assessed in future studies. Classical signs and symptoms of TMD were not good predictors, including pain intensity at rest. However, the fact that pain on chewing pre-treatment was found to be predictor may be explained by the fact that patients with higher levels of affective disorders, as in the case of nrTMD, perceive their problems as more severe, despite the absence of visible signs of dysfunction (Freeman et al., 1998).

In addition, we were also able to show that the neuropsychological and psychosocial profile between irritable bowel syndrome (IBS) and nrTMD patients were much more similar than those between the IBS and the rTMD group. On average, both nrTMD and IBS patients did worse in the neuropsychological tests, with the exception of the reaction time tests. In addition, the former patient groups showed significantly higher levels of sleep disorders, depression, fatigue and lower levels of energy level than the latter.

Taken in combination, the data suggested that neuropsychological differences exist between the nrTMD versus rTMD patients. In addition, it was also demonstrated that two unrelated chronic pain conditions (i.e. IBS and nrTMD) may share similar
neuropsychological characteristics. Therefore, solid and reproducible evidence was provided with the use of neuropsychological tests not only in favor of the biopsychosocial model of chronic pain, but also in favor of the concept that these two unrelated chronic pain conditions may be part of a similar chronic pain syndrome with obvious implications in disease classification and treatment modalities (Moldofsky, 1993a). Our study strongly supports multidisciplinary management of both nrTMD and IBS populations.
### IX. APPENDICES

#### APPENDIX I: RECRUITMENT LOG OF ALL NEW OROFACIAL PAIN PATIENTS

<table>
<thead>
<tr>
<th>Diagnostic Categories</th>
<th>Number of new cases</th>
<th>Proportion of new cases (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporomandibular joint pain</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>without muscle involvement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients who did not show for appointment</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Neurological disorders</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Headaches:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tension</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Cluster</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Migraine</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Temporomandibular Disorders</td>
<td>22*</td>
<td>19*</td>
</tr>
<tr>
<td>Psychological disorders</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Trauma</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Motor Vehicle Accidents</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Trigeminal Neuralgia</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Atypical Facial Pain</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Fibromyalgia</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Dental pain (endodontics)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Oral pathology diseases</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Post-surgical pain</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Medical legal cases</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Loss of posterior support</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>116</td>
<td>100</td>
</tr>
</tbody>
</table>

Source: Mount Sinai Hospital Orofacial Pain Clinic between June 01 and August 13, 1996

* Recruitment rate for TMD patients who met our inclusion criteria = 50%
### APPENDIX II: NUMBER AND PROPORTION OF MISSING VALUES PER VARIABLE

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>VALID CASES</th>
<th>MISSING CASES</th>
<th>MISSING CASES (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SIMPLE REACTION TIME TEST</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2. MULTIPLE CHOICE REACTION TIME TEST</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3. MCRT WITH CONFLICT</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4. MCRT WITH CONSTRAINT</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5. CVLT - CORRECT RESPONSES</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6. CVLT - CLUSTERS</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7. CVLT - PERSEVERATIONS</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8. CVLT - INTRUSIONS</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9. CONSONANT TRIGRAM TEST</td>
<td>98</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10. SLEEP ASSESSMENT QUESTIONNAIRE</td>
<td>99</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11. BECK DEPRESSION INVENTORY</td>
<td>98</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>12. FATIGUE</td>
<td>98</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>13. ENERGY LEVEL</td>
<td>98</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14. EDUCATIONAL LEVEL</td>
<td>99</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15. EMPLOYMENT</td>
<td>99</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16. INCOME</td>
<td>97</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>17. MAXIMUM MOUTH OPENING</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>18. OVERTBITE</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>19. OVERTET</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>20. PERCUSSION SENSITIVITY</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>21. CARIES</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>22. EXACERBATION</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>23. TMJ FACIAL</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>24. TMJ EAM</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>25. TMJ SOUNDS</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>26. MASSETER</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>27. TEMPORALIS</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>28. STERNOCLEIDOMASTOID</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>29. MEDIAL PTERYGOID</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>30. LATERAL PTERYGOID</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>31. INSERTION TEMPORALIS</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>32. TOTAL JOINT SCORE</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>33. TOTAL MUSCLE SCORE</td>
<td>58</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>34. PAIN AT REST PRE-TREATMENT</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>35. PAIN AT REST POST-TREATMENT</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>36. PAIN ON CHEWING PRE-TREATMENT</td>
<td>59</td>
<td>1</td>
<td>1.7</td>
</tr>
<tr>
<td>37. PAIN ON CHEWING POST-TREATMENT</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>38. OBJECTIVE IMPROVEMENT</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>39. SUBJECTIVE IMPROVEMENT</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>40. AGE</td>
<td>90</td>
<td>10</td>
<td>10*</td>
</tr>
<tr>
<td>41. TREATMENTS</td>
<td>55</td>
<td>5</td>
<td>9*</td>
</tr>
<tr>
<td>42. LENGTH OF TREATMENT</td>
<td>55</td>
<td>5</td>
<td>9*</td>
</tr>
<tr>
<td>43. CLINICIAN</td>
<td>55</td>
<td>5</td>
<td>9*</td>
</tr>
<tr>
<td>44. IRRITABLE BOWEL SYNDROME</td>
<td>54</td>
<td>6</td>
<td>11.1*</td>
</tr>
<tr>
<td>45. PAIN DURATION</td>
<td>51</td>
<td>9</td>
<td>17.6*</td>
</tr>
</tbody>
</table>

* Proportion of missing values exceeding 5% of total valid cases
### Appendix III: Associations Between Variables with Missing Values Exceeding 5% of Valid Cases Versus Responding (Group I) and Non-Responding TMD (Group II)

<table>
<thead>
<tr>
<th>Independent Variables:</th>
<th>Responding TMD, Group I (n=36)</th>
<th>Non-responding TMD, Group II (n=24)</th>
<th>Irritable Bowel Syndrome Group III (n=20)</th>
<th>Non-Pain Population Group IV (n=15)</th>
<th>Test of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of treatments (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion of valid cases</td>
<td>91.7</td>
<td>91.7</td>
<td>88.9</td>
<td>91.7</td>
<td>p = 1.0 §; NS</td>
</tr>
<tr>
<td>proportion of missing values</td>
<td>8.3</td>
<td>8.3</td>
<td>11.1</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td><strong>Length of treatment (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion of valid cases</td>
<td>91.7</td>
<td>91.7</td>
<td>88.9</td>
<td>91.7</td>
<td>p = 1.0 §; NS</td>
</tr>
<tr>
<td>proportion of missing values</td>
<td>8.3</td>
<td>8.3</td>
<td>11.1</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td><strong>Treating clinician (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion of valid cases</td>
<td>91.7</td>
<td>91.7</td>
<td>88.9</td>
<td>91.7</td>
<td>p = 1.0 §; NS</td>
</tr>
<tr>
<td>proportion of missing values</td>
<td>8.3</td>
<td>8.3</td>
<td>11.1</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td><strong>Irritable Bowel Syndrome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- comorbidity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion of valid cases</td>
<td>86.1</td>
<td>83.3</td>
<td>80.6</td>
<td>86.1</td>
<td>p = 0.7 §; NS</td>
</tr>
<tr>
<td>proportion of missing values</td>
<td>13.9</td>
<td>16.7</td>
<td>19.4</td>
<td>13.9</td>
<td></td>
</tr>
<tr>
<td><strong>Pain duration (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proportion of valid cases</td>
<td>83.3</td>
<td>87.5</td>
<td>95.0</td>
<td>100.0</td>
<td>p = 0.2 §; NS</td>
</tr>
<tr>
<td>proportion of missing values</td>
<td>16.7</td>
<td>12.5</td>
<td>5.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

* p<0.05, ** p<0.01, ***p<0.001

§ Chi Square test for differences between proportions
APPENDIX IV: CORRELATIONS BETWEEN MULTIPLE CONFOUNDING VARIABLES IN
THE CVLTCR (GROUP I VERSUS II)

<table>
<thead>
<tr>
<th>Correlations:</th>
<th>BDI</th>
<th>SAQ</th>
<th>FATIGUE</th>
<th>ENLEVEL</th>
<th>AGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVLTCR</td>
<td>.1105</td>
<td>-1.099</td>
<td>-.0293</td>
<td>.1394</td>
<td>-.1663</td>
</tr>
<tr>
<td>CVLTCL</td>
<td>.0510</td>
<td>.0661</td>
<td>.0994</td>
<td>-.1219</td>
<td>.0550</td>
</tr>
<tr>
<td>CCC</td>
<td>.0334</td>
<td>.0701</td>
<td>.1434</td>
<td>-.1187</td>
<td>.3646</td>
</tr>
<tr>
<td>BDI</td>
<td>1.0000</td>
<td>.1073</td>
<td>.2947</td>
<td>-.2804</td>
<td>.2733</td>
</tr>
<tr>
<td>SAQ</td>
<td>.1073</td>
<td>1.0000</td>
<td>.4346</td>
<td>-.0462</td>
<td>.1824</td>
</tr>
<tr>
<td>FATIGUE</td>
<td>.2947</td>
<td>.4346</td>
<td>1.0000</td>
<td>-.5385*</td>
<td>-.0735</td>
</tr>
<tr>
<td>ENLEVEL</td>
<td>-.2804</td>
<td>-.0462</td>
<td>-.5385*</td>
<td>1.0000</td>
<td>-.1163</td>
</tr>
<tr>
<td>AGE</td>
<td>.2733</td>
<td>.1824</td>
<td>-.0735</td>
<td>-.1163</td>
<td>1.0000</td>
</tr>
<tr>
<td>EDLEVEL</td>
<td>-.2756</td>
<td>.0619</td>
<td>-.3373</td>
<td>.1883</td>
<td>.5176*</td>
</tr>
<tr>
<td>INCOM</td>
<td>-.1589</td>
<td>.2758</td>
<td>-.1481</td>
<td>.0201</td>
<td>.3397</td>
</tr>
<tr>
<td>PRESTPR</td>
<td>.0869</td>
<td>.1809</td>
<td>.1726</td>
<td>-.1706</td>
<td>.1348</td>
</tr>
<tr>
<td>PCHEWPRE</td>
<td>-.2095</td>
<td>.0554</td>
<td>.3872</td>
<td>-.1523</td>
<td>-.5207*</td>
</tr>
<tr>
<td>EMPL</td>
<td>-.4163</td>
<td>-.3834</td>
<td>.0050</td>
<td>-.2750</td>
<td>-.3657</td>
</tr>
</tbody>
</table>

N of cases: 21  1-tailed Signif: * .01  ** .001

".
" is printed if a coefficient cannot be computed

---

Correlations: EDLEVEL INCOM PRESTPR PCHEWPRE EMPL

<table>
<thead>
<tr>
<th>Correlations:</th>
<th>EDLEVEL</th>
<th>INCOM</th>
<th>PRESTPR</th>
<th>PCHEWPRE</th>
<th>EMPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVLTCR</td>
<td>.2551</td>
<td>.0422</td>
<td>-.0623</td>
<td>-.0293</td>
<td>-.0580</td>
</tr>
<tr>
<td>CVLTCL</td>
<td>.4498</td>
<td>.2677</td>
<td>.1084</td>
<td>-.1615</td>
<td>-.2203</td>
</tr>
<tr>
<td>CCC</td>
<td>.6117*</td>
<td>.6132*</td>
<td>-.3723</td>
<td>-.5033</td>
<td>-.4312</td>
</tr>
<tr>
<td>BDI</td>
<td>-.2756</td>
<td>-.1589</td>
<td>.0869</td>
<td>-.2095</td>
<td>-.4163</td>
</tr>
<tr>
<td>SAQ</td>
<td>.0619</td>
<td>.2758</td>
<td>.1809</td>
<td>.0554</td>
<td>-.3834</td>
</tr>
<tr>
<td>FATIGUE</td>
<td>-.3373</td>
<td>-.1481</td>
<td>.1726</td>
<td>.3872</td>
<td>.0050</td>
</tr>
<tr>
<td>ENLEVEL</td>
<td>.1883</td>
<td>.0201</td>
<td>-.1706</td>
<td>-.1523</td>
<td>-.2750</td>
</tr>
<tr>
<td>AGE</td>
<td>.5176*</td>
<td>.3397</td>
<td>.1348</td>
<td>-.5207*</td>
<td>-.3657</td>
</tr>
<tr>
<td>EDLEVEL</td>
<td>1.0000</td>
<td>.5069*</td>
<td>-.2553</td>
<td>.4416</td>
<td>-.2692</td>
</tr>
<tr>
<td>INCOM</td>
<td>.5069*</td>
<td>1.0000</td>
<td>-.3027</td>
<td>-.5541*</td>
<td>-.3007</td>
</tr>
<tr>
<td>PRESTPR</td>
<td>-.2553</td>
<td>-.3027</td>
<td>1.0000</td>
<td>.3795</td>
<td>.1708</td>
</tr>
<tr>
<td>PCHEWPRE</td>
<td>-.4416</td>
<td>-.5541*</td>
<td>.3795</td>
<td>1.0000</td>
<td>.5879*</td>
</tr>
<tr>
<td>EMPL</td>
<td>-.2692</td>
<td>-.3007</td>
<td>.1708</td>
<td>.5879*</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

N of cases: 21  1-tailed Signif: * .01  ** .001

".
" is printed if a coefficient cannot be computed
APPENDIX V: GLOSSARY OF ABBREVIATIONS USED IN THIS STUDY

AC: Adaptive Copers TMD patients
ACC: Anterior Cingulate Cortex
AD: Alzheimer’s Disease
ADA: American Dental Association
ANOVA: Analysis of Variance
BCPP: Borg’s Category Scale for Ratings of Perceived Pain
BDI: Beck Depression Inventory
BRS: Behaviour Rating Scale
CBT: Cognitive Behavioral Therapy
CCC: Brown-Peterson Consonant Trigram Auditory Memory Task
CESF: Clinical Examination Short Form
CHI: Closed Head Injury
CMI: Cornell Medical Index
CMI: Craniomandibular Index
CNS: Central Nervous System
CVLT: California Verbal Learning Test
CVLTCL: California Verbal Learning Test - Clusters
CVLTCR: California Verbal Learning Test - Correct Responses
CVLTI: California Verbal Learning Test - Intrusions
CVLTP: California Verbal Learning Test - Perseverations

DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders version III - Revised

DSM-III: Diagnostic and Statistical Manual of Mental Disorders version III

DSM-IV: Diagnostic and Statistical Manual of Mental Disorders version IV

DYS: Dysfunctional TMD patients

EDLEVEL: Educational Level

EMPL: Employment

ENERGY: Energy Level

EPI: Eysenck Personality Inventory

FEPs: Fatigue Energy Level Pain Self-Completing Questionnaire

FM: Fibromyalgia

GI: Gastrointestinal Disorders

HAS: Hamilton Anxiety Scale

HDS: Hamilton Depression Scale

IBS: Irritable Bowel Syndrome

ICC: Intraclass Correlation Coefficient

ID: Interpersonally Distressed TMD patients

IMPATH:TMJ: Microcomputer Assessment of Behavioral and Psychosocial Factors in Craniomandibular Disorders

INCOM: Income Level

IQ: Intelligence Coefficient

iTMD: idiopathic Temporomandibular Disorders
LENGTX: Length of Treatment

M-VAS: Mechanical Visual Analogue Scale

MANOVA: Multiple Analysis of Variance

MCRT: Multiple Choice Reaction Time test

MCRTCF: Multiple Choice Reaction Time test with Conflict

MCRTCT: Multiple Choice Reaction Time test with Constraint

MMO: Maximum Mouth Opening

MMPI: Minnesota Multiphasic Personality Inventory

MPD: Myofascial Pain Dysfunction

MPI: the Multidimensional Pain Inventory

MPQ: McGill Pain Questionnaire

MRI: Magnetic Resonance Imaging

MVA: Motor Vehicle Accident

NPV: Negative Predictive Value

NREM: Non-Rapid Eye Movement

NSAIDs: Non-Steroid Anti-inflammatory Drugs

NWC: Number of Words Chosen

OB: Overbite

OJ: Overjet

OR: Odds Ratio

PCHEWPRE: Pain on Chewing Pre-treatment

PD: Parkinson’s Disease
PDQ: Pilowsky’s Depression Questionnaire
PPV: Positive Predictive Value
PRESTPRE: Pain at Rest Pre-treatment
PRS: Pain Relief Scale
pTMD: post-traumatic Temporomandibular Disorders
RC: Reliability of Change
RDC/TMD: Research Diagnostic Criteria for Temporomandibular Disorders
REM: Rapid Eye Movement
SAQ: the University of Toronto Sleep Assessment Questionnaire
SCL-90: Symptom Checklist-90
SCL-90R: Symptom Checklist-90 Revised
SD: Standard Deviation
SPSS PC+: Statistical Package for the Social Sciences for Personal Computer Plus
SRT: Simple Reaction Time test
SSI: Symptom Severity Index
TBI: Traumatic Brain Injury
TENS: Transcutaneous Electric Stimulation
TMAI: Taylor Manifest Anxiety Inventory
TMD: Temporomandibular Disorders
TMJ: Temporomandibular Joint
TXS: Number of Treatments Provided
VAS: Visual Analogue Scale
VDS: Verbal Descriptor Scale

VPS: Verbal Pain Scale

VS: Verbal Scale

VSRT: Verbal Selective Reminding Test

WHLCS: Wallston's Health Locus of Control Scale
X. BIBLIOGRAPHY


CESTA A, MOLDOFSKY H, SAMMUT C: The University of Toronto Sleep Assessment Questionnaire (SAQ). *Sleep Research* 1996;25:486.


DWORKIN SF, VON KORFF MR, LERESCHE L Multiple pains and psychiatric disturbance: an epidemiologic investigation. *Arch Gen Psychiatry* 1990d:47;239-44.


HUDSON JI, GOLDENBERG DL, POPE HG Jr, KECK PE Jr, SCHLESINGER L:
Comorbidity of fibromyalgia with medical and psychiatric disorders. *Am J Med*


JACOBSON NS, FOLLETTE WC, REVENSTORF D: Psychotherapy outcome research:
methods for reporting variability and evaluating clinical significance. *Behav Ther*

KARNI A *et al.*: Dependence on REM sleep of overnight improvement of a perceptual

KATZMAN R, ARONSON M, FULD P, KAWAS C, BROWN T, *et al.*: Development of


KLEINKNECHT RA *et al.*: Correspondence between subjective report of
temporomandibular disorder symptoms and clinical findings. *JADA*
1986;113:257-261.


KOPP S: Constancy of clinical signs in patients with mandibular dysfunction. *Community

KOPP S, WENNEBERG B. Intra- and inter-observer variability in the assessment of


TAYLOR W: How to use guide to sample size calculation. McMaster University, 1981.


TURK DC, RUDY TE, SORKIN B: Neglected topics in chronic pain treatment outcome studies: determination of success. Pain 1993a;53:3-16.


WILSON L, DWORINK SF, WHITNEY CW, LERESCHE L: Somatization and pain

WILSON MA, MCNAUGHTON BL: Reactivation of hippocampal ensemble memories

1990 criteria for classification of fibromyalgia: Report of the multicenter criteria

WOODFORDE JM, MERSKEY H: Some relationships between subjective measures of

ZARB GA et al.: Temporomandibular Joint and Masticatory Muscle Disorder.

YOSHTAKE H: Three Characteristic Patterns of Subjective Fatigue Symptoms.