Reviewer's report

Title: Cognitive performance in relapsing remitting multiple sclerosis patients during 2 years treatment with intramuscular interferon-beta-1a: an observational study in daily practice using a brief computerized cognitive battery

Version: 5 Date: 11 April 2011

Reviewer: Richard Benedict

Reviewer's report:

I am pleased to review this manuscript again in revised form. The authors are to be commended for subjecting their work to peer review, which is critical for new computerized testing batteries to be widely accepted in the neurological and psychological literatures. Unfortunately, I continue to have some major concerns with the manuscript in its present form.

Abstract

Explain that Power of Attention etc are subtests from the CDR and what they measure.

Explain who are the controls, this is very important. Otherwise the d values are meaningless.

Stable is not defined, do you mean that the mean values did not change, good test-retest reliability?

Introduction

“Cognitive dysfunction has a significant impact on social functioning and health-related quality of life (HR-QoL).” Requires reference, actually several studies have shown that self-report QOL measures are more related to depression than NP measures.

The authors are stubbornly holding on to the idea that this is a treatment effects study as they state “The present study investigated cognition during 2 years treatment with IM INF#-1a in RRMS patients, treated in daily practice.”

This sentence is still problematic: “Three of the MACFIMS tests (D-KEFS sorting, 10/36 and especially BVMTR) are dependent on motor responding, in a patient group where this may be impaired.” First the 10/36 is not in the MACFIMS.

Second the DKEFS requires no more motor function than does using a mouse or keyboard. BVMTR has a copy trial to control for manual coordination effects.

The authors seem to be promoting (selling) the CDR and the first author is employed by the company. In the Competing Interests section these relationships need to be explicated.

The issue of alternate forms requires elucidation. The CDR randomly generates alternate forms using a computer algorithm as is common with computerized
batteries. However, that does not mean that the forms are equivalent in difficulty – this is an untested assumption inherent in the program. It could also hamper test-retest reliability, which should be acknowledged later.

Methods and Results

I believe the authors wish to state that oral responses are written down by the examiner.

In addition, there is no evidence that visual recall is assessed. Do the subjects describe figures and the examiner writes down these responses also.

Please explain how the subtest scores are calculated.

What does (training) mean on the time point measures?

All of the Pearson r reliability coefficients should be reported, that is all time points. See for example Benedict RHB, Duquin JA, Jurgensen S, Rudick R, Feiticher J, Munschauer F, Panzara MA, Weinstock-Guttman B (2008). Repeated Assessment of Neuropsychological Deficits in Multiple Sclerosis using the Symbol Digit Modalities Test and the MS Neuropsychological Screening Questionnaire. Multiple Sclerosis, 14:940-946

I remain very concerned about the selection of controls from an industry developed database. Much more explanation is needed here. I suspect that this is not a control group per se, but rather a standardization sample which is routinely employed by the users of the CDR. This is tantamount to a study on the WAIS in MS and comparing patient performance to the manual norms. This is not necessarily a fatal flaw, but the authors should be up front about this and fully describe these control data.

For all validity coefficients, the time point needs to be clear. I suggest that the authors stick to correlating the baseline values and perhaps the termination values.

Discussion

The Discussion should be shortened and the authors should stick much more to their data.

The authors state: In this study, clinically active RRMS patients who started IM INFb-1a treatment in daily practice showed stable cognition, both in terms of change over time and proportion of patients with cognitive impairment. This is a veiled intimation that the drug was effective, and it is repeated in the conclusions. ALL such references should be removed, as there is no control group.

The CDR System is the most widely used computerized assessment of cognition in clinical trials. Citation needed or remove the sentence.

Again in the Discussion there is exaggeration of the merits of CDR, suggesting bias among the authors. The CDR does not cover the same domains as the MACFIMS [eg learning over multiple trials, visual recall memory, abstract reasoning, verbal fluency]. Later, the authors’ descriptions of the BRB and
MACFIMS memory tests are not correct. There are several errors here and I will not list them – I suggest the authors consult the manuals or more detailed descriptions of these tests.

The conclusions regarding sensitivity and reliability of CDR may hold, depending on what is found with the recommended analyses above. I am especially interested in the reliability data from the other time points not yet mentioned.