

Reliability and clinical utility of a Portuguese version of the Abnormal Involuntary Movements Scale (AIMS) for tardive dyskinesia in Brazilian patients

H. Tonelli¹, D. Tonelli¹,
G.R. Poiani²,
M.A.B.F. Vital²
and R. Andreatini²

¹Psiquiatria da Aliança Saúde, Pontifícia Universidade Católica do Paraná,
Santa Casa, Curitiba, PR, Brasil
²Departamento de Farmacologia, Universidade Federal do Paraná, Curitiba,
PR, Brasil

Abstract

The objective of the present study was to evaluate the reliability and clinical utility of a Portuguese version of the Abnormal Involuntary Movements Scale (AIMS). Videotaped interviews with 16 psychiatric inpatients treated with antipsychotic drugs for at least 5 years were evaluated. Reliability was assessed by the intraclass correlation coefficient (ICC) between three raters, two with and one without clinical training in psychopathology. Clinical utility was assessed by the difference between the scores of patients with (N = 11) and without (N = 5) tardive dyskinesia (TD). Patients with TD exhibited a higher severity of global evaluation by the AIMS (sum of scores: 4.2 ± 0.9 vs 0.4 ± 0.2 ; score on item 8: 2.3 ± 0.3 vs 0.4 ± 0.2 , TD vs controls). The ICC for the global evaluation was fair between the two skilled raters (0.58-0.62) and poor between these raters and the rater without clinical experience (0.05-0.29). Thus, we concluded that the Portuguese version of the AIMS shows an acceptable inter-rater reliability, but only between clinically skilled raters, and that it is clinically useful.

Key words

- Reliability
- Scale
- Tardive dyskinesia
- Clinical utility

Correspondence

M.A.B.F. Vital
Laboratório de Fisiologia e
Farmacologia, SNC, Centro
Politécnico
Departamento de Farmacologia,
UFPR
81531-990 Curitiba, PR
Brasil
Fax: +55-41-266-2042
E-mail: vital@bio.ufpr.br

Received July 10, 2002
Accepted December 20, 2002

The introduction of effective medications for the treatment of schizophrenia and other psychoses was a major advance in twentieth-century medicine (1). These drugs were initially named “neuroleptics” because of their tendency to produce acute extrapyramidal side effects, and some clinicians and investigators felt this was an essential characteristic of a drug that would have therapeutic activity in the treatment of schizophrenia. Since that time the belief in the association between

“neuroleptic” effect and clinical antipsychotic efficacy has been largely, though not completely, abandoned (1).

A few years after the introduction of antipsychotic drugs, Schönecker described the occurrence of tardive dyskinesia (TD) as an involuntary movement disorder characterized by a variable mixture of the following features: orofacial and lingual dyskinesia, tics, grimacing, truncal or axial muscle involvement, chorea, athetosis, and dystonias

(2). Speech and respiration may also be affected (1-3). TD usually persists for months after the neuroleptic is discontinued and may be irreversible. Since the early description of this syndrome in the late 1950's, there have been numerous prevalence surveys in various populations (2,4,5). Prevalence estimates have varied enormously and a variety of methodologic problems have made it difficult to establish estimates of true prevalence with any degree of reliability (2). For example, in inpatients from Salvador, Brazil, the average prevalence of TD was found to be 1.65% (6), while in ambulatory schizophrenic patients from Rio de Janeiro, Brazil, it was 26.3% (7).

Many scales are used to evaluate acute extrapyramidal effects induced by neuroleptics such as the Simpson and Angus Scale (8), Akathisia Scale (9), Extrapyramidal Symptom Rating Scale (10), and St. Hans Scale (11). For TD evaluation, the best known scales are the Abnormal Involuntary Movements Scale (AIMS) (12) and the Tardive Dyskinesia Rating Scale (13).

The AIMS, proposed by Guy (12), is one of the instruments most frequently used to assess TD in various populations. The scale consists of 10 items including symptoms and attitudes, whose intensity varies from 0 to 4. This instrument was translated into Portuguese (14), but the translated version has not yet been validated. Moreover, instructions for evaluation and a scale version with anchored scores have been proposed recently (15).

The usefulness of clinical scales is limited by their psychometric properties. In particular, reliability, which is the measure of the stability and the errors of measurement, is a basic feature of any instrument. Reliability can be measured between raters (inter-rater reliability) and between different times for the same rater (test-retest reliability).

The objective of the present study was to evaluate the inter-rater reliability and the diagnostic discrimination (clinical utility) of

the Portuguese version of the AIMS in Brazilian patients.

All subjects were native Brazilian Portuguese speakers. A total of 16 inpatients with at least 5 years of antipsychotic treatment were evaluated. Eleven patients were also diagnosed with TD by the Diagnostic and Statistical Manual IV (DSM-IV) diagnostic criteria (16). The remaining five patients without TD composed the control group.

The AIMS is a 10-item scale administered by an observer that rates the items from 0 = absent to 4 = severe. The total score was obtained as the sum of scores for all items. The patients were contacted and, after the explanation of the aim and the procedures of the protocol, were invited to participate in the study. After acceptance, they (or an immediate family member or a legal guardian) signed a term of informed consent. The protocol was approved by the Ethics Committee of Hospital de Clínicas, Universidade Federal do Paraná.

The patients were submitted to a videotaped interview by one researcher (HT) who is a clinical psychiatrist and who conducted all interviews. The interviews were then analyzed by two other raters, i.e., a clinical psychiatrist (DT) and a basic psychopharmacologist (MABFV) with minimal training in clinical skills. The clinical psychiatrists had at least nine years of clinical psychiatric experience. The inter-rater reliability was measured by pairing two raters in three different manners, i.e., pairs consisting of one of the experienced clinical raters (HT or DT) and the inexperienced clinical rater (MABFV) and one consisting of the two experienced raters (HT and DT). The mean scores obtained by a diagnostic blind rater (DT) were compared by the Student *t*-test, with the level of significance set at $P < 0.05$. The reliability of the parameters was evaluated by the intraclass correlation coefficient (ICC) (17). After calculation, the reliability indexes were classified as poor (< 0.50), fair (0.50-0.75) or good (> 0.75) according to

Spitzer and Endicott (18).

There are two methods to evaluate global severity of TD by AIMS: the sum of items and the highest score for any item. Figure 1 shows these scores for TD patients and controls. There was a significant difference between groups both in AIMS total score ($t = 2.97, P < 0.02$) and in item 8 ($t = 3.72; P < 0.01$).

The reliability measures are shown in Table 1. For global severity, reliability between trained raters and naive raters was poor (ICC between 0.05 and 0.29). The reliability between two experienced raters was classified as fair (ICC between 0.58 and 0.62).

There are no reliable or well-validated strategies for identifying a true case of TD, and a variety of medical conditions may produce abnormal involuntary movements that may be difficult to distinguish from TD. There continues to be some controversy as to the amount of antipsychotic drug treatment necessary or sufficient to produce abnormal involuntary movements in some psychiatric patients. In the present study we found that TD patients had higher scores than psychiatric patients without TD. Although the scores of TD seem low, it can be noted that the minimum score criterion for the diagnosis of TD in some studies was a total AIMS score of 3 (19). This suggests that the Portuguese version of the AIMS is clinically useful to discriminate between patients with and without TD.

The inter-rater reliability was fair between two raters with previous clinical psychopathological training. These results could be viewed as an indication of the need for a more extensive training on this scale, despite the use of specific guidelines for rating the severity provided by the AIMS version employed in the present study. We first discussed each item after an evaluation of three videorecorded interviews of patients with TD symptoms and then discussed the measurements. Reliability was found to be poor when the rater was not trained in clinical

skills. The need for clinical experience for raters to use a scale has been emphasized several times, and for the AIMS in particular (19). However, some investigators have suggested that the AIMS may be employed by non-physician persons with training, so that its use could be incorporated into routine practice (15). Our results clearly do not endorse this position, but support the need for experienced clinical raters. Overall, the ICC observed in the present study was lower than observed with the English version of the AIMS. For example, Lane et al. (19) found an ICC of 0.79-0.86 for experienced raters for total score or severity and Gerlach et al. (11) found an ICC of 0.60 to 0.72. One possible explanation for this difference is that in the present study we included patients with and without TD in contrast with other studies that only evaluated patients with TD (11,19). In the former case, a higher frequency of zero score was expected, which

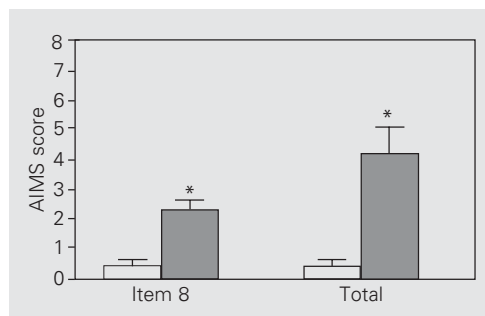


Figure 1. Scores obtained with the Portuguese version of the Abnormal Involuntary Movements Scale (AIMS) for control patients (open bars, N = 5) or tardive dyskinesia patients (closed bars, N = 11). Data are reported as means \pm SEM scores obtained by one rater who was unaware of the diagnosis. * $P < 0.05$ compared to control (Student *t*-test).

Table 1. Inter-rater reliability for total score (item 8 or sum of items) of the Portuguese version of the AIMS.

	ICC		
	1 x 2 ^a	2 x 3 ^a	1 x 3 ^a
Item 8	0.14	0.62	0.10
Sum of items	0.29	0.58	0.05

Reliability was evaluated by the intraclass correlation coefficient (ICC).

^aPairs of raters: 1 = rater without psychopathological training, 2 and 3 = raters with psychopathological training.

could reduce the reliability coefficient (20). Since experience with TD influences the reliability of the AIMS (11,19), another hypothesis for the lower ICC found in our study could be a greater exposure to patients with TD of the investigators involved in the other studies, which are usually conducted by persons working in movement disorder clinics. One factor that might have influenced our results was the fact that one rater (HT) knew the diagnosis of each patient, in fact, he selected the patients for the videotape record. This is the reason for our decision to compare the AIMS score between TD patients and controls using only the experienced rater (DT). With respect to reliability, we think that this aspect did not bias the ICC in a significant manner since it

measures the similarity of the rating between raters and not the discrimination between patients.

Taken together, the present results suggest that the psychometric properties of the Portuguese version of the AIMS could be viewed as satisfactory when used by raters with clinical training, although extensive training may be needed, and that the use of the scale for Brazilian patients is acceptable. However, the present results should be considered with caution since they may have been influenced by the small size of the sample studied. Thus, more research is needed to further explore the psychometric properties of the Portuguese version of the AIMS, although its clinical utility appears to be established.

References

- Kane JM (1995). Tardive dyskinesia: epidemiological and clinical presentation. In: Bloom FE & Kupfer DJ (Editors), *Psychopharmacology: The Fourth Generation of Progress*. Raven Press, Ltd., New York, 1485-1495.
- Bassit DP & Louzã Neto MR (1999). *Discinesia Tardia*. Casa do Psicólogo, São Paulo, SP, Brazil.
- Andrade LA, Bertolucci PHF & Pereira JS (1984). Discinesia tardia: I. Fisiologia e tratamento. *Arquivos de Neuro-Psiquiatria*, 42: 362-370.
- Jeste DV & Wyatt RJ (1981). Changing epidemiology of tardive dyskinesia: an overview. *American Journal of Psychiatry*, 138: 297-309.
- Kane JM & Smith JM (1982). Tardive dyskinesia: prevalence and risk factors, 1959 to 1979. *Archives of General Psychiatry*, 39: 473-481.
- Andrade Filho AS, Santos SRS, Medeiros CL, Carneiro CC, Figueiroa FLS, Silva MC & Dunningham WA (1994). Prevalência de discinesia tardia em pacientes internados nos hospitais psiquiátricos de Salvador, Bahia, Brasil. *Arquivos de Neuro-Psiquiatria*, 52: 354-357.
- Piedade RAM & Nardi AE (1988). Discinesia tardia: estudo epidemiológico em pacientes esquizofrênicos ambulatoriais. *Jornal Brasileiro de Psiquiatria*, 37: 81-86.
- Simpson GM & Angus JWS (1970). A rating scale for extrapyramidal side effects. *Acta Psychiatrica Scandinavica*, 212 (Suppl): 11-19.
- Barnes TRE (1989). A rating scale for drug-induced akathisia. *British Journal of Psychiatry*, 154: 672-676.
- Chouinard G, Ross-Chouinard A, Annabie L & Jones BD (1980). Extrapyramidal symptom rating scale. *Canadian Journal of Neurological Sciences*, 7: 233 (Abstract).
- Gerlach J, Kosgaard S, Clemmesen P, Lund Lauersen AML, Magelund G, Noring U, Povlsen UJ, Bech P & Casey DE (1993). The St. Hans rating scale for extrapyramidal syndromes: reliability and validity. *Acta Psychiatrica Scandinavica*, 87: 244-252.
- Guy W (1976). *ECDEU Assessment Manual for Psychopharmacology*. US Department of Health, Education and Welfare, Washington, DC.
- Simpson GM, Lee JH, Zoubock B & Gardos G (1979). A rating scale for tardive dyskinesia. *Psychopharmacology*, 64: 171-179.
- Louzã Neto MR (1998). Escalas de avaliação de efeitos colaterais de antipsicóticos. *Revista de Psiquiatria Clínica*, 25 (Edição especial): 357-367.
- Munetz MR & Benjamin S (1988). How to examine patients using the abnormal involuntary movements scale. *Hospital and Community Psychiatry*, 39: 1172-1177.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*. 4th edn. American Psychiatric Association, Washington, DC.
- Kramer MS & Feinstein AR (1981). Clinical biostatistics. LIV. The biostatistics of concordance. *Clinical Pharmacology and Therapeutics*, 29: 111-123.
- Spitzer RL & Endicott J (1980). Psychiatric rating scales. In: Kaplan H, Freeman A & Sadock B (Editors), *Comprehensive Textbook of Psychiatry*. 3rd edn. Williams & Wilkins, New York, 2015-2031.
- Lane RD, Glazer WM, Hansen TE, Berman WH & Kramer SI (1985). Assessment of tardive dyskinesia using the abnormal involuntary movement scale. *Journal of Nervous and Mental Disease*, 173: 353-357.
- Bartko J & Carpenter N (1976). On the methods and theory of reliability. *Journal of Nervous and Mental Disease*, 163: 307-317.