The Schedule for the Assessment of Drug-Induced Movement Disorders (SADIMoD)

Loonen, Anton J.M.; Doorschot, Cees H.; Van Hemert, Dianne A.; Oostelbos, Marc C.J.M.; Sijben, Arnold E.S.

Published in:
International Journal of Neuropsychopharmacology

DOI:
10.1017/s1461145701002589

Publication date:
2001

International Journal of Neuropsychopharmacology (2001), **4**, 347–360. Copyright © 2001 CINP
DOI: 10.1017/S1461145701002589

ARTICLE

The Schedule for the Assessment of Drug-Induced Movement Disorders (SADIMoD): inter-rater reliability and construct validity

Anton J. M. Loonen[1], Cees H. Doorschot[2], Dianne A. van Hemert[3], Marc C. J. M. Oostelbos[4], Arnold E. S. Sijben[5] and the MASEAS Team*

[1] *Delta Psychiatric Hospital, PO Box 800, NL-3170 DZ Poortugaal, The Netherlands*
[2] *GGZ 's-Hertogenbosch, PO Box 10.150, NL-5260 GB Vught, The Netherlands*
[3] *Tilburg University, Department of Psychology, PO Box 90.153, NL-5000 LE, Tilburg, The Netherlands*
[4] *Two Way bv, Muntelbolwerk 1, NL-5213 SZ 's-Hertogenbosch, The Netherlands*
[5] *ABC/Opes bv, Zutphensestraatweg 17, NL-6881 WN Velp, The Netherlands*

Abstract

The Schedule for the Assessment of Drug-Induced Movement Disorders (SADIMoD) is a newly developed instrument, consisting of a compilation of rating scales, to measure the severity of drug-induced movement disorders: dystonia, dyskinesia, Parkinsonism, akathisia, ataxia, and several types of tremors. The inter-rater reliability and the construct validity of this scale were investigated. Six investigator teams were trained by means of a standard package of instruction material to such an extent that a single member of the team could represent the entire team. Thirty-one patients [20 male/11 female; 57.1 ± 6.5 yr (mean ± S.D.)] with a variety of movement disorders were recorded on videotape according to the SADIMoD Schedule. Single representatives of all six teams scored these video recordings. To this set the existing SADIMoD ratings of 82 patients were added to form the so-called 'total data set'. These patients were examined by 6 different researchers, who rated 4, 8, 10, 14, 18 and 28 patients, respectively, mostly in the context of a research protocol. A specific subset consisted of 12 patients that were examined three times with a two-weekly interval without any change of their medical condition or treatment. The 6 ratings of the 31 individual patients correlated to a highly significant degree, with Kendall's Coefficients of Concordance of 0.436 to 0.891 (median 0.717). The same was true for the 6 ratings of the 7 SADIMoD subscales (median 0.578, range 0.462–0.715) Considering the total data set, the homogeneity of the various subscales was good (Cronbach's $\alpha =$ 0.81–0.94, median: 0.87). The SADIMoD dyskinesia and dystonia subscales showed a highly significant mutual correlation. The Parkinsonism subscale correlated highly significantly with the rest and postural tremor subscales and to a lesser extent with the akathisia and ataxia subscales. An analysis of variance showed that the three ratings in the subset of 12 patients were not significantly different for any scale. Also Scheffé tests for homogenous subsets did not reveal any significant differences. When investigated under 'real world' circumstances, the inter-rater reliability of the SADIMoD was found to be satisfying. The instruction material, that was developed and used in this study, fully comes up to the requirements. The construct validity of the SADIMoD is more than sufficient.

Received 21 December 2000; Reviewed 20 March 2000; Revised 30 April 2001; Accepted 6 May 2001

Key words: Ataxia, basal ganglia diseases, movement disorders – drug-induced, rating scales.

Address for correspondence: Dr A. J. M. Loonen, Delta Psychiatric Hospital, PO Box 800, NL-3170 DZ Poortugaal, The Netherlands.
Tel.: +31 10 503 13 13 *Fax*: +31 10 503 13 88 *E-mail*: Anton.Loonen@wxs.nl

* The MASEAS Team consists of the examiners and/or raters and the camera-operators making the video-recordings of the six participating centres. **Delta Psychiatric Hospital, Poortugaal**: B. Resida, rater; M. Kwakkelstein, rater; C. Sterrenburg, rater; L. Timmerman, investigator; Th. van Zonneveld, rater; R. Keulemans, assistant. **GGZ 's-Hertogenbosch OP; 's-Hertogenbosch**: E. Delescen, investigator; J. van der Veer, rater; W. Garenfeld, rater; G. Ackermans, assistant. **GGZ 's-Hertogenbosch VP, Vught**: C. Doorschot, investigator; W. van Rooij, rater; Ch. Maton, rater; H. Kaethoven, rater/assistant. **GGZ Midden-Brabant, Tilburg**: W. Standaert, investigator; M. Pepping, assistant. **GGZ Oost-Brabant, Rosmalen**: B. Jogems, investigator; F. Schenkelaars, rater; M. van Gool, rater/assistant; J. Sturkenboom, rater/assistant. **PCA Zuid/Nieuw-West, Amsterdam**: L. Perquin, investigator; R. Halkema, rater; J. Wesselius, rater; K. Hanekroot, rater; E. Beyaert, rater; P. Kempe, assistant.

Introduction

The SADIMoD is a newly developed instrument to measure the severity of a variety of syndromes (Doorschot et al., 1993, 1994; Loonen et al., 1993, 2000). It is intended to be used in clinical trials on psychotropic drugs, in trials on the therapy of drug-induced movement disorders and in daily practice to examine the long-term course of movement disorders in individual patients.

Instruments to measure drug-induced movement disorders

Several types of instruments have been examined for their suitability to assess drug-induced movement disorders in clinical trials (de Leon and Simpson, 1992; Gardos et al., 1977; Loonen et al., 1997). These instruments can be broadly divided into instrumentational techniques, frequency counting techniques and rating scales. Jogems-Kosterman et al. (1998) have recently reported on a very promising instrumentational assessment technique using the quantitative characteristics of the pen movements during the execution of writing and drawing tasks. This was shown to be a very elegant method to assess bradykinesia and postural tremors, but is still in an experimental stage. The utilization of many other, often very sophisticated, instrumental techniques to measure movement disorders have been described in the literature (de Leon and Simpson, 1992; Gardos et al., 1977). However, these methods are generally only suitable to quantify one or only a few aspects of the possible movement disorders, e.g. the severity of rigidity (Caligiuri et al., 1989; Chan et al., 1987; Teräväinen et al., 1989), or the frequency and amplitude of a resting hand tremor (Sinnaeve, 1989). Moreover, special equipment, computer programs and trained personnel are necessary to collect and analyse the movement data. These techniques are,

Table 1. Examples of rating scales to assess movement disorders

Akathisia
Barnes Rating Scale for Drug-Induced Akathisia (Barnes, 1989)
Hillside Akathisia Scale (Fleischhacker et al., 1989)
Prince Henry Hospital Akathisia Rating Scale (Sachdev, 1994)
Dyskinesia
Abnormal Involuntary Movement Scale (Guy, 1976)
Dyskinesia Identification System Condensed User Scale (Kalachnik and Sprague, 1993)
Dyskinesia Check List (Crane et al., 1969)
Tardive Dyskinesia Rating Scale of Escobar (Reda et al., 1975)
Tardive Dyskinesia Rating Scale of Simpson et al. (1979)
Abbreviated Dyskinesia Rating Scale of Simpson et al. (1979)
Dystonia
Fahn–Marsden Dystonia Movement Scale (Burke et al., 1985)
Parkinsonism[a]
Echelle pour Bilan Extra-Pyramidal (Bordeleau et al., 1967).
Scale for the Clinical Assessment of Parkinsonism (Mindham, 1976).
Parkinson's Disease Rating Scale of Webster (1968).
Simpson-Angus Rating Scale for Extrapyramidal Side Effects (Simpson and Angus, 1970).
Short Parkinson's Evaluation Scale (Rabey et al., 1997).
Targeting Abnormal Kinetic Effects (Doller Wojcik et al., 1980).
Unified Parkinson's Disease Rating Scale (Fahn et al., 1987).
Multiple disorders
Dokumentationssystem von Degkwitz, Heinrich und Hippius (Hippius and Logemann, 1970).
Proforma for recording clinical data on motor disorders (Hershon et al., 1972; Kennedy et al., 1971).
Extrapyramidal Symptom Rating Scale (Klett and Caffey, 1972).
Rating Scale for Reversible Extrapyramidal Symptoms (Chien et al., 1974).
Extrapyramidal Symptom Rating Scale (Chouinard et al., 1984).
Involuntary Movement and Extrapyramidal Side-Effects Scale of May and Van Putten (Marder et al., 1984).
Smith-TRIMS Tardive Dyskinesia Scale (Smith et al., 1977, 1983).
Sct. Hans Rating Scale for Extrapyramidal Syndromes (Gerlach, 1979; Gerlach and Korsgaard, 1983).

[a] Scales intended to measure genuine Parkinson's disease are not (extensively) covered.

therefore, seldom suitable to be applied in ordinary clinical trials or in daily practice on a routine basis. More appropriate for the latter purpose are the so-called rating scales. In a variety of global or multi-item rating scales it is attempted to measure specific movement disorders by quantifying clinical observations under standardized conditions. Some of these scales are developed to assess only one type of movement disorder (de Leon and Simpson, 1992; Gardos et al., 1977; Loonen and Zwanikken, 1987). In performing clinical trials, combinations of such scales are applied to measure the separate movement disorders. Other scales intend to quantify a set of disorders separately and simultaneously (Table 1). These scales can be applied as the only instrument to measure that set of disorders. Generally speaking, rating scales have a better validity than the more objective techniques, but their reliability is less certain. The reliability of the assessment by means of rating scales can be improved by giving precise instructions on how to examine the patient, when to score the individual movement disorders and how to rate the severity of their distinct components. Moreover, by videotaping the standardized examination, the movement disorders can be scored by several independent raters at every suitable moment.

Characteristics of the most important scales

For research purposes, it is nowadays common practice to use a combination of rating scales such as the Simpson and Angus (1970) Rating Scale for Extrapyramidal Side Effects (SEE), the Barnes (1989) Rating Scale for Drug-Induced Akathisia (BAS) and the Abnormal Involuntary Movement Scale (AIMS) of the National Institute of Mental Health (NIMH) (Guy, 1976).

The AIMS is a rating scale for dyskinesias devised by the Psychopharmacology Research Branch of the NIMH (Gardos et al., 1977). This scale consists of separate ratings on a 5-point scale of dyskinesias of the face, lips, jaw, tongue, arms, legs and trunk. In addition, three global severity ratings of abnormal movements are added: as seen by the observer, the patient's reaction to them, and the incapacitation that results from them. Two additional items deal with the dental status. An examination schedule is described in detail. A differentiation is made between spontaneous and activated movements. The rating of dyskinesias that occur during activity should be reduced with 1 point in order to obtain the final score. This procedure probably decreases the reliability of this scale. Moreover, the validity of the items on the dental status is doubtful, and the usefulness of the global ratings can be disputed. Therefore, some authors only consider the first 7 or 8 items of the scale (Loonen et al., 1992). At least eight reports exist on studies concerning the inter-rater reliability of the AIMS, which are summarized by Edson et al. (1997). Six of them use the sum of the scores for items 1–7 as an outcome. In all but two studies the investigators use live subject interviews instead of videotapes. Some authors expressed the inter-rater reliability as Pearson's r and others by interclass correlation coefficients. The results are summarized in Table 2. As can be concluded from the Bergen et al. (1988) study, the inter-rater reliability of the separate items is rather poor. The inter-rater reliability of the total 1–7 score is far better, especially when the raters are intensively trained as a single group (Sweet et al., 1993). However, the AIMS is probably not suitable to assess dyskinesias in long-term trials without videotaping the patients. Tracy et al. (1997) have observed, that the inter-rater reliability decreased from 0.87 to 0.60 within a few months without periodic joint training.

The SEE, which was published by Simpson and Angus in 1970, was actually the second edition of a scale published in 1964 (Simpson et al., 1964). The original scale was expanded to 10 items rated on a 5-point scale, including tremor and salivation. The evaluation of the trunk muscles was dropped from the original scale, as it was too difficult to quantify. An important point of criticism on the SEE can be, that 6 or 7 (depending on the question whether or not the glabella tap reflects something else) of the 10 items, one way or another way deal with rigidity. Moreover, the instructions on how to examine the patient are quite brief and sometimes unclear. For example, different examiners were found to execute and interpret the glabella tap in a deviant manner and the author's advice was sought as to examination and interpretation of results. Frequently, a patient goes from a rating of 0 to one of 4 without there ever appearing to be an intermediate score (G. M. Simpson, personal communication). In addition, two items of the scale – head dropping and leg pendulousness – suppose that an examination table is readily available. As this is often not the case the examiner is obliged to change the examination procedure. These imperfections probably decrease the inter-rater reliability to a great extent. Although the authors have modified the scale in order to correct this, the original scale is still in use. Apart from the original publication of Simpson and Angus (1970) only one paper was found, that addressed the estimation of the inter-rater reliability of the SEE. Sweet et al. (1993) had 10 patients assessed by 4 raters and calculated an intraclass correlation coefficient (ICC) of 0.79 for the total score. For the separate items the ICC varied from 0.33 to 1.00. Simpson and Angus (1970) had the inter-rater reliability of two doctors assessed by comparing their scorings of 14 patients and had found a correlation coefficient of 0.87. However, these results may be flattered as the raters are

Table 2. Inter-rater reliability studies for the AIMS

Authors	Sample size	Raters	Inter-rater reliabilty
Chien et al. (1977)	$n = 38$	$n = 2$	Pearson's $r = 0.87$
Smith et al. (1979a)	$n = 39–48$	$n = 2$	Pearson's $r = 0.87$ (total 1–7)[a]
Smith et al. (1979b)	$n = 64–80$	$n = 2$	Pearson's $r = 0.87$ (total 1–7)[b]
Lane et al. (1985)	$n = 33$	$n = 2–4$	Pearson's $r = 0.81$ (total 1–7)[c] ICC = 0.79 (total 1–7)
Gerlach et al. (1993)	$n = 30$	$n = 7$	ICC = 0.60 – 0.71 (total 1–7)[d]
Sweet et al. (1993)	$n = 10$	$n = 4$	ICC = 0.91 (total 1–7)[e]
Bergen et al. (1988)	$n = 30$	$n = 2$	ICC = 0.30–0.63 (items 1–8)[f]
Tracy et al. (1997)	$n = 19$	$n = 18$	ICC = 0.75 (total 1–7)[g]

[a] Six pairs formed by four individual raters examined a total of 377 in-patients. The inter-rater reliability was expressed as the mean Pearson correlation averaged over the six rater team combinations using Fisher's Z transformation. The combined inter-rater reliability of the separate items varied from $r = 0.52$ to $r = 0.82$.

[b] Three pairs formed by three individual raters examined a total of 213 outpatients. The combined inter-rater reliability of the separate items varied from 0.48 to 0.99.

[c] Different combinations of 2–4 raters were formed by four psychiatrists to examine 33 outpatients. Average Pearson correlations for more than two raters were computed using Z score transformations. The ICC was calculated according to Bartko and Carpenter (1976). The combined inter-rater reliability of the separate items varied from 0.46 to 0.80.

[d] Seven raters scored 30 in-patients three times with an interval of 2 wk. The ICC was calculated according to Bartko and Carpenter (1976). Two experienced raters performed far better (ICC = 0.76–0.87) than three inexperienced ones (ICC = 0.53–0.72).

[e] The ICCs were calculated using the General Linear Models Procedure of the SAS Release 6.06.

[f] Inter-rater reliability of separate items expressed as the mean of the unweighted Cohen's κ over 5 ratings of 30 subjects by two raters.

[g] A selection of a total of 29 investigators from 9 different centres scored videotapes of 19 patients. The intraclass correlation coefficient was calculated according to Fleiss (1986). The ICC for the separate items varied from 0.18 to 0.74.

probably not sufficiently independent. Therefore, these results are better not considered.

The BAS was derived from an examination of the signs and symptoms shown by 104 consecutively admitted acute psychiatric in-patients who had received antipsychotic drugs and 89 chronic psychiatric outpatients on long-term antipsychotic medication (Barnes, 1989, 1992). This instrument comprises ratings on a 4-point scale of the observable characteristic restless movements, of the patient's awareness of this restlessness and of the patient's distress related to it. In addition, a global severity rating on a 6-point scale is present, with clear definitions of each scale point. This global assessment offers the opportunity to distinguish pseudo-akathisia. Brief instructions are given on how to examine the patient. Barnes and colleagues studied the inter-rater reliability of two raters in a sample of 42 drug-treated schizophrenic in-patients (Barnes, 1989, 1992). The inter-rater reliability, expressed as linearly weighted Cohen's κ, ranged from 0.74 to 0.95 for the 4 rating scale items. Sweet et al. (1993) estimated an ICC of 0.93 for the total score and ICCs varying from 0.83 to 0.94 for the 4 items. Edson et al. (1997) measured an ICC of 0.73 for the total score after having the videotapes of 10 subjects scored by 9 different raters. However, in the latter case the videotape material may have been less suitable to assess patients with the BAS. It may be concluded that, in general, the BAS performs very well.

The Extrapyramidal Symptom Rating Scale (ESRS), developed by the psychiatrist-pharmacologist G. Chouinard and the neurologist A. Ross-Chouinard in 1979, has to our knowledge never been published in a scientific journal or book. The validity and reliability are said to have been studied, but the results are not described in the manual of the scale. Nevertheless, the significance of this scale has shown an immense increase after its utilization in several large international trials on new antipsychotic drugs (Simpson and Lindenmayer, 1997). The scale includes a subjective questionnaire for Parkinsonian, dystonic and dyskinetic symptoms. Unfortunately, precise instructions are missing as to how the examiner should interpret the answers to these questions when rating the patient's condition. A good point is the inclusion of a standard examination procedure, which includes the observation of the patient executing a set of well-described tasks. The manual gives an indication of what disorders to look for during the execution of these tasks. The ESRS consists of multi-item ratings on a 7-point

scale for Parkinsonism, acute torsion dystonia, non-acute or chronic torsion dystonia and dyskinetic movements. Clinical global impressions of severity on a 9-point scale are added for dyskinesia, Parkinsonism and dystonia, as well as a rating of the stage of Parkinsonism according to Hoehn and Yahr (1967). In the Parkinsonism subscale the rigidity of every limb is counted separately and tremors are scored in as much as 8 body areas apart. On the other hand akathisia is scored as a symptom of Parkinsonism, which is quite peculiar in our opinion. Moreover, the distinction that is made between acute and non-acute torsion dystonia does not increase the transparency. The examiner will seldom observe acute torsion dyskinesia himself and the description is quite detailed. The most important methodological objection is the procedure to rate tremors and dyskinetic movements. These movements are rated along two axes or dimensions, i.e. the amplitude of the movement and the frequency of their occurrence. It is absolutely unclear whether or not this is a valid procedure to measure the severity of these movement disorders. In addition, the rules for the calculation of factors and total scores are rather complex. Six factors were identified on the basis of the assessment of 305 neuroleptic-treated chronic schizophrenic outpatients by a single investigator that accounted for 67.1% of the variance in the items of the scale: hypokinetic Parkinsonism, orofacial dyskinesia, trunk/limb dyskinesia, akathisia, tremor and tardive dystonia.

The Sct. Hans Rating Scale (SHRS) is a multidimensional scale developed in the 1970s. A preliminary version of the hyperkinesia subscale was taken as a reference by Chien et al. (1977). Gerlach published the final version in 1979. The scale consists of four subscales for hyperkinesia, Parkinsonism, dystonia and akathisia (Gerlach and Korsgaard, 1983; Gerlach et al., 1993). The hyperkinesia scale scores dyskinesias in 8 topographical regions. Furthermore, a global score is included. The movements are scored on a 0- to 6-point scale when the patient is sitting and relaxed (passive phase) and when the patient is actively involved in doing something, e.g. writing or walking. The Parkinsonian subscale consists of 8 items, which are – like everywhere else in the SHRS – rated on a 7-point scale. In addition, a global score is added. A global score only represents dystonia. The Akathisia subscale consists of a global score of subjective and one of objective symptoms. Moreover, the severity of sedation, depression and anxiety are scored. The scale has been used in conjunction with a video-recording and a standardized examination procedure has been developed. Originally, the SHRS included a special section for the detailed analysis of oral dyskinetic movements, but this part is only useful in specialized studies. Seven raters in 30 psychiatric in-patients with tardive dyskinesia tested the inter-rater reliability of the Hyperkinesia and Parkinsonism subscales (Gerlach et al., 1993). A slightly modified AIMS was used for comparison (see Table 2). The inter-rater and test–retest reliability was generally high for the experienced raters (ICC = 0.82–0.98), but were considerably lower for less experienced ones. Convergent validity was found between the dyskinesia scales and the AIMS and divergent validity between the other scales. The Parkinsonism subscale had a high construct validity and the Dyskinesia subscale had not.

The Unified Parkinson's Disease Rating Scale (UPDRS) was developed by the UPDRS Development Committee in 1984 (Fahn et al., 1987) and is nowadays the most widely used scale in clinical research and drug trials of patients with Parkinson's disease. It includes an evaluation of self-reported disability [activities of daily living (ADL) section] and a clinical scoring by a physician [motor examination (ME) section]. In addition, the mental function, the presence and extent of dyskinesias and clinical fluctuations are evaluated. Finally, the presence of other complications of therapy is assessed. The UPDRS is very suitable to assess the impairments and disabilities of a patient with Parkinson's disease, but this makes it a less appropriate instrument to measure Parkinsonism in an acute psychotic patient. Therefore, this scale is seldom applied in clinical trials on psychotropic drugs.

The SADIMoD

The SADIMoD offers researchers the possibility to use one instrument instead of a combination of rating scales such as the AIMS, SEE and BAS. Moreover, the SADIMoD allows a unified, rather in-depth evaluation of movement dysfunctions that may not necessarily stem from the same pathophysiological substrates. In comparison to other composite scales such as the ESRS or the SHRS, the SADIMoD has a more clear structure, assesses more disorders separately and independently and contains more transparent definitions for the assessable disorders and severity items. The instrument consists of subscales to quantify the severity of dyskinesias (7 items, passive phase and active phase), dystonia (9 items), Parkinsonism (8 items), akathisia (2 items), several types of tremors (3 items), ataxia (5 items), and 4 mental symptoms. Moreover, each subscale contains a total score and a global score, the latter offering the examiner the opportunity to express his personal opinion concerning the nature and severity of the disorder. When the SADIMoD was developed, the SHRS for Extrapyramidal Symptoms served as a framework (Gerlach et al., 1979, 1983). Several well-known movement scales were added (Loonen et al., 1997). The definitions for the severity scores (on a 5-point

scale) were either derived from these movement scales or adopted from the UKU Side Effect Assessment Scale (Lingjærde et al., 1987). In order to create a coherent whole, the item definitions and subscales were adjusted to each other. The only exception is the subscale for measuring dystonia, that was derived from the Fahn–Marsden Dystonia Scale and in which the original structure was preserved (Burke et al., 1985). A standardized examination schedule and video registration protocol have been developed in order to be able to elicit and measure the movement disorders under fixed conditions (Loonen et al., 1994). A sample of the handwriting is taken in order to detect micrographia, tremors or ataxia (Haase, 1977). This is stored to serve for future comparison. In addition, the patient is questioned concerning subjective feelings and past events. So, the final package consists of an examination schedule and instruction for videotaping, a questionnaire, a writing test, a rating form and a glossary. This glossary contains a description and classification of the movement disorders as well as item definitions and severity scores of the individual subscales (Loonen et al., 1997). In order to be able to complete the score form of the SADIMoD, the patient is videotaped while being submitted to a strictly standardized examination schedule. Thereafter, further information is inquired verbally. The whole procedure takes about 25 min, whereas about 14 min are taped. The final assessment takes about 30 min per video registration.

In an earlier report (Loonen et al., 2000) the test–retest variability was described. Even when assessed with a time interval of 110.3 ± 58.0 d, the test and retest ratings correlated to a highly significant degree with Spearman's correlation coefficients of 0.57 to 0.88 (median 0.69). Convergent validity was found between the SADIMoD dyskinesia and (to a lesser extent) dystonia scales and the AIMS as well as between the akathisia subscales and the BAS, with divergent validity with the other subscales. The SEE discriminated less well between the Parkinsonism subscale and the other subscales.

Using the results of an open trial of the acute effects of sertindole in acutely admitted psychotic patients, the concurrent validity with the BAS, AIMS, Fahn–Marsden Dystonia Movement Scale, and Webster's Parkinson Disease Rating Scale (WPDRS) was assessed. Moreover, it was examined whether the SADIMoD is more sensitive to change. The highest Spearman correlation coefficients (0.88–0.96; highly significant $p < 0.01$) were found for ratings on the SADIMoD subscales and their corresponding scales. Comparing the change of the scores after the initiation of treatment, the SADIMoD showed a similar or significantly larger (vs. BAS, WPDRS) sensitivity than the comparators (A. J. M. Loonen et al., unpublished observations).

Aim of the present study

Although the SADIMoD was already used in two studies (Jogems-Kosterman et al., 1998; Loonen et al., 1999), in which it showed to be a very sensitive instrument to measure changes of the severity of for example Parkinsonism, formal studies to demonstrate its instrumental nomological characteristics were still lacking. Therefore, we undertook a multi-centre study in which we investigated the concurrent validity of the SADIMoD with the AIMS, the SEE and the BAS as well as the test–retest reliability (Loonen et al., 2000). In the same study we also investigated the inter-rater reliability. Moreover, by adding these data to the filed ratings of patients in earlier studies and daily practice, we had now access to the data of over 100 SADIMoD ratings, which offered the opportunity to examine the construct validity of our scale. In the present paper, we report on the results of these analyses.

Experimental procedures

The design of the SADIMoD validation study and characteristics of the examined patient population have been described elsewhere in detail (Loonen et al., 2000). Six investigators from different centres each formed an investigator team. The investigator received the SADIMoD manual, the prescribed examination materials and an instruction video. He was instructed to use this material in order to train his team to an acceptable standard and to achieve a consensus, thereby allowing him to consider all individual team members capable of representing the team. Thereafter, he selected up to 6 psychiatric patients, who suffered from at least one mildly severe, relatively stable, movement disorder that was possibly, probably, or certainly psychotropic drug induced. The same examiner evaluated each patient in one session by means of four different assessment instruments. These instruments were: the SADIMoD, the Simpson–Angus (1970) Rating Scale for Extrapyramidal Side Effects (SEE), the Barnes (1989) Rating Scale for Drug-Induced Akathisia (BAS), and the Abnormal Involuntary Movement Scale (AIMS) of the NIMH (Guy, 1976). After having examined all patients, the complete team scored the video recordings of these patients. This resulted in the first ratings of the patients of that centre. After all six participating centres had completed the rating of their patients, another six sessions were organized to reassess the videotapes. During these sessions single representatives of each team came together in one of the centres and scored the patients of that centre. The raters could ask for a reshow of fragments of the videotape, but no case discussions were allowed between the representatives of

the different teams. This resulted in the second ratings of these patients, that served to estimate the inter-rater variability. This study was undertaken in accordance with the Declaration of Helsinki and the Dutch legislation concerning the performance of medical research and in particular videotaping psychiatric patients. The protocol was submitted and approved by the Medical Ethical Committee 'Toetsingscommissie Zuid-Nederland.'

In order to create the data set for the determination of the construct validity only the first ratings of the above-mentioned patients were used. To this set the existing SADIMoD ratings of 82 patients were added. Twenty-eight of these patients had been examined and scored by the investigator himself (A.J.M.L.), the others were examined by 5 different researchers who were all personally trained by the investigator (A.J.M.L.). In various cases the investigator participated in rating the videotapes. Only patients that were scored from a standard videotape were considered. One of the researchers examined only 4 patients, the others respectively 8, 14, 10 and 18. When the same patient had been examined more than once, only the first examination was used for the data set. The original Dutch version of the SADIMoD was used to score the patients. An unchanged translation of this manual was used in the SADIMoD validation study. Most patients were scored in the context of a research protocol (Jogems-Kosterman et al., 1998; Loonen et al., 1993, 1994, 1999; Van Eindhoven et al., 1995). A specific subset of these existing examinations consisted of a group of 13 patients who were recorded more than once with a two-weekly interval, without any change to their medical condition or treatment (Van Eindhoven et al., 1995). From 12 patients three recordings existed and of 1 patient only two recordings were made. Each patient was assessed in one scoring session by the same rater. This data was used to calculate the variability introduced by the examination procedure and also reflects day to day variability under natural conditions. The results are compared with the test–retest and inter-rater variability of the current validation study.

Statistical analysis

Analyses were performed with SPSS for Windows 9.01.

Inter-rater reliability

A data set with only second ratings was used (see Results section for a description). Nine Kendall's Coefficients of Concordance were calculated concerning the six different ratings of all individual total subscale scores in the 31 patients. Moreover, 31 Kendall's W were calculated of the six different ratings of the total subscale scoring profile for each individual patient. This coefficient tests the sufficiency with which different raters similarly score the same patient. This agreement should be highly significant ($p < 0.001$).

Internal consistency

The total data set was used to establish the internal consistency. The phenomenon was measured by calculating Cronbach's α coefficient (Cronbach, 1951). This coefficient tests the sufficiency with which one item can substitute for the other. A Cronbach's α coefficient of 0.70 or higher is usually considered acceptable (Nunnally, 1970).

Concurrent validity

Spearman correlation coefficients were calculated for all scales to express the degree of convergent (or divergent) validity. These analyses were performed on the total data set. Convergent validity is expected between scales that measure the same clinical phenomenon or that measure phenomena that have the same dose–response relationship.

Factor structure

An exploratory factor analysis was performed on the total data set. Factor loadings were rotated according to Varimax with Kaiser normalization. The identified factors are expected to correspond to the scales that measure the distinctive clinical phenomena.

Variability of ratings

A smaller subset (to be described in the results section) was used to establish the variability of scores across ratings for 12 patients. One-way ANOVA and the Scheffé test for homogeneity of subsets were conducted. In addition, Spearman correlation coefficients were calculated for the three ratings of all scales to be able to compare this variability with the test-retest and inter-rater variability. In order to allow a power analysis when the SADIMoD is used in doing clinical research, the essential characteristics are given.

Results

Data sets

Three data sets were used. The first data set consisted of all second ratings of all patients of the multi-centre

validation study (Loonen et al., 2000). This data set was used to calculate the inter-rater variability. The second data set, hereafter named the total data set, consisted of SADIMoD data for 82 patients from the pilot study, combined with data for 31 patients from the validation study (Loonen et al., 2000). Only first ratings were included. This data set was used to estimate the construct validity. The third data set consisted of three ratings for 12 patients. Time between ratings varied from 7 to 11 d. First ratings for these patients were also included in the total data set.

Missing data

All data were screened for irregularities before analysis. Some missing data could be inferred from an available total score; these were added. Remaining missing data in the total data set included scores of 4 patients for the global dystonia scale, of 1 patient for the global Parkinsonism scale, and of 1 patient for the postural tremor scale. Further, for 11 patients data were missing on all four psychic symptoms. All missing data were included as missing data in subsequent analyses. One case showed 83.33% missing data and was removed. This brought the number of patients in the total data set to 112. The third data set did not contain any missing data.

Patients

From the total data set of 112 patients, 20 patients suffered of dystonia, 52 of active and 40 of passive dyskinesia, 62 of Parkinsonism, 31 of akathisia, and 44 of ataxia of at least mild severity all according to the Physician's Global Impression as reflected by a global score of 2 or more on the SADIMoD subscales. Further, 17 patients suffered of tremor at rest, 35 of postural tremor, and 24 of intention tremor of at least mild severity (score $\geqslant 2$). Thirty-five patients suffered from Parkinsonism combined with either active, passive or both dyskinesias. Active and passive dyskinesia occurred together in 37 patients. Fourteen patients showed dystonia as well as ataxia. Three patients scored 2 or more on all global scales. Seven patients scored 2 or more on all global scales but one (being the Dystonia scale in 5 cases, the Ataxia scale in 1 case, and the Akathisia scale in 1 case). Seventy-two patients suffered from more than one movement disorder simultaneously.

Inter-rater reliability

The inter-rater variability of the scorings of the different SADIMoD subscales by the six different raters is shown in Table 3. For all subscales the ratings corresponded with each other to a highly significant degree. When the variability of the scorings of the six raters was considered concerning their scorings of the 9 total subscales for the 31 individual patients, the inter-rater reliability was even better. The 31 Kendall's W varied from 0.436 to 0.891 with a median value of 0.717. In all but 2 cases (with $p < 0.01$), the correlation was highly significant ($p < 0.001$).

Table 3. Inter-rater reliability expressed as Kendall's Coefficients of Concordance of all SADIMoD total subscale scores ($n = 31$)

Subscale of the SADIMoD	Kendall's W
Dystonia	0.462**
Parkinsonism	0.643**
Dyskinesia, passive phase	0.715**
Dyskinesia, active phase	0.708**
Ataxia	0.685**
Tremor at rest	0.526**
Postural tremor	0.486**
Intention tremor	0.477**
Akathisia	0.578**

* $p < 0.01$; ** $p < 0.001$.

Table 4. Construct validity expressed as Cronbach's α coefficient for the subscales of the SADIMoD

Subscale of the SADIMoD	Cronbach's α
Dystonia (9)[a]	0.83
Dystonia including global score	0.88
Parkinsonism (8)	0.82
Parkinsonism including global score	0.87
Dyskinesia passive phase (7)	0.82
Dyskinesia passive including global score	0.87
Dyskinesia active phase (7)	0.83
Dyskinesia active including global score	0.88
Dyskinesia total (14)	0.91
Dyskinesia total including global score	0.93
Ataxia (5)	0.84
Ataxia including global score	0.89
Akathisia (2)	0.62
Akathisia including global score	0.81

[a] Number of items in scale are within parentheses.

Internal consistency

Cronbach's α coefficients for the individual subscales of the SADIMoD are shown in Table 4. This coefficient was calculated with and without taking the global score into consideration. All but one coefficients were higher than 0.70. As the number of items influences the homogeneity of the subscales, Spearman–Brown's formula for the

Table 5. Spearman correlations between the SADIMoD scales for the total data sets ($n = 112$)

Subscale of the SADIMoD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Dystonia total	1														
(2) Dystonia global	0.84**	1													
(3) Parkinsonism total	0.17	0.21*	1												
(4) Parkinsonism global	0.14	0.16	0.92**	1											
(5) Dyskinesia passive	0.44**	0.38**	0.13	0.12	1										
(6) Dyskinesia passive global	0.42**	0.41**	0.16	0.15	0.93**	1									
(7) Dyskinesia active	0.43**	0.38**	0.13	0.12	0.88**	0.80**	1								
(8) Dyskinesia active global	0.43**	0.40**	0.20*	0.17	0.84**	0.84**	0.92**								
(9) Ataxia total	0.30**	0.35**	0.24*	0.25**	0.31**	0.33**	0.28**	0.30**	1						
(10) Ataxia global	0.31**	0.40**	0.30**	0.32**	0.29**	0.33**	0.23**	0.29**	0.89**	1					
(11) Tremor at rest	−0.06	−0.06	0.30**	0.28**	0.08	0.09	0.05	0.06	0.10	0.10	1				
(12) Postural tremor	0.19	0.18	0.31**	0.36**	0.22*	0.19	0.28**	0.31**	0.09	0.10	0.25**	1			
(13) Intention tremor	0.11	0.21*	−0.03	0.02	0.14	0.07	0.08	0.08	0.15	0.16	0.05	0.40**	1		
(14) Akathisia total	0.28**	0.23*	0.22*	0.24*	0.42**	0.41**	0.39**	0.40**	0.17	0.14	−0.01	0.20*	−0.09	1	
(15) Akathisia global	0.25**	0.19*	0.15	0.17	0.37**	0.35**	0.34**	0.35**	0.18	0.12	0.01	0.13	−0.16	0.89**	1

* $p < 0.05$; ** $p < 0.01$.

reliability of a lengthened test (Lord and Novick, 1968) was used to establish corrected reliabilities of the relatively short subscales Akathisia (2 items) and Akathisia including global score (3 items). When the Akathisia subscale would have consisted of 7 items (the average number of items in the various subscales of the SADIMoD), Cronbach's α would have measured 0.85. This coefficient would be 0.91 for a 7-item Akathisia subscale including Global score.

Concurrent validity

Three data sets have been analysed with respect to the mutual correlation between the scoring of the various subscales: the data of the 31 patients of the validation study, the data of the 81 patients of the pilot study, and their combination in the total data set. As only small differences existed between these data sets, the findings with the total data set adequately reflect the characteristics of the schedule (Table 5). The highest correlation appeared to exist between the Dystonia and Dyskinesia subscale scores and between the Akathisia and Dyskinesia subscale scores. A somewhat smaller correlation coefficient was found between the Dystonia and the Ataxia or Akathisia subscale scores as well as between the Dyskinesia and the Ataxia subscale scores. In addition, the active phase Dyskinesia subscale scoring correlated to a similar degree with the postural tremor rating. On the other hand, the Parkinsonism subscale ratings showed a low correlation with the Dystonia subscale and all but one of the Dyskinesia subscales. The scoring of the Parkinsonism subscale correlated best with the rest tremor, postural tremor, and ataxia scorings, and to a lesser extent with the akathisia scorings.

Factor structure

The 15 total and global scales in the total data set were factor-analysed. Exploratory factor-analysis resulted in six factors (eigenvalues 5.112, 2.287, 1.902, 1.464, 1.180 and 1.019), explaining a total of 86.43% of variance. The factor solution was Varimax rotated. Factor I can be regarded as the Dyskinesia factor, as the active and passive Dyskinesia scales (both total and global) loaded highest on this factor. The second factor was named the Parkinsonism factor. The total and global Parkinsonism scales loaded highest on this factor, as well as tremor at rest and postural tremor. Factor III can be considered the Ataxia factor, with the total and global ataxia scales loading highly. The total and global Akathisia scales loads highest on the fourth factor. The fifth factor was named the Dystonia factor, as both the total and global Dystonia scales loaded highest on this factor. Finally, the sixth factor could be regarded as the Tremor factor, as both

Table 6. Variability of the ratings of 12 clinically stable patients who were examined three times without a change of medication had occurred

SADIMoD subscale	Mean of 3 ratings in 12 patients	Standard deviation of these 36 ratings	Standard error of the differences between the 3 ratings
Dystonia	2.528	3.325	1.348
Dystonia global	0.583	0.692	0.281
Parkinsonism	7.972	6.162	2.540
Parkinsonism global	1.222	1.045	0.433
Dyskinesia passive	4.528	3.722	1.523
Dyskinesia passive global	1.083	0.937	0.393
Dyskinesia active	5.750	3.341	1.393
Dyskinesia active global	1.389	0.838	0.340
Ataxia	3.972	2.667	1.108
Ataxia global	1.028	0.736	0.304
Tremor at rest	0.750	1.360	0.571
Postural tremor	0.944	0.715	0.294
Intention tremor	0.528	0.609	0.252
Akathisia	3.250	2.740	1.134
Akathisia global	1.444	1.157	0.474

Table 7. Spearman correlation between the three ratings of 12 clinically stable patients who were examined thrice without a change of treatment had occurred

	Ratings 1 and 2	Ratings 2 and 3	Ratings 1 and 3
Dystonia	−0.24	0.42	−0.06
Dystonia global	−0.06	0.44	−0.06
Parkinsonism	0.92**	0.89**	0.80**
Parkinsonism global	0.70*	0.76**	0.64*
Dyskinesia passive	0.67*	0.63*	0.81**
Dyskinesia passive global	0.45	0.31	0.58*
Dyskinesia active	0.75**	0.83**	0.69*
Dyskinesia active global	0.63*	0.32	0.13
Ataxia	0.72**	0.67*	0.82**
Ataxia global	0.67*	0.63*	0.60*
Tremor at rest	1.00**	0.61*	0.61*
Postural tremor	0.21	0.46	0.58*
Intention tremor	0.38	0.69*	0.54
Akathisia	0.75**	0.89**	0.68*
Akathisia global	0.69*	0.97**	0.64*

* $p < 0.05$; ** $p < 0.01$.

intention tremor and postural tremor loaded highest on this factor.

Intra-individual variability

Table 6 shows the standard errors of the differences between the first, second and third ratings of the 12 patients from the third data set. In addition, the mean score and standard deviation is given. No significant differences were observed. In Table 7 the correlation coefficients between these three ratings are given. Generally, the global ratings correlated to a somewhat lesser extent than the total scores. The three ratings of Dystonia, Postural tremor and Intention tremor scales showed no or only a limited correlation.

Variability of ratings

An analysis of variance (ANOVA) was conducted on the 15 total and global scales in the third data set, to establish

whether the three ratings of the 12 patients were similar. For no scale were the three ratings found to be significantly different. Also, Scheffé tests for homogeneous subsets did not reveal any significant differences in means between the three ratings.

Discussion

In the present article we describe some important nomological characteristics of the SADIMoD. This schedule is certainly not the first instrument that has been developed to measure drug-induced extrapyramidal movement disorders. However, the SADIMoD is unique in also quantifying other than the classical 'extrapyramidal' movement disorders such as ataxia and tremors. In addition, the nomological characteristics of most rating scales have not been established. This is even true when these scales have been used as a prime instrument to measure the advantages of a new antipsychotic drug, such as the Extrapyramidal Symptom Rating Scale (ESRS) of Chouinard et al. (1984). When investigated, most rating scales appear to perform so poorly, that they 'deserve a decent burial' and subsequent fall into oblivion (Cunningham Owens, 1999). To such scales belong the very often applied Rating Scale for Extrapyramidal Side Effects (SEE) of Simpson and Angus (1970) and the AIMS of the NIMH (Guy, 1976). It is very unfortunate that, for example, the AIMS has been the principle instrument to establish the effects of up to 2 yr of treatment with d-vitamin E in 150 subjects with tardive dyskinesia in a recently published well-designed, large-scale, 9-centre long-term clinical trial (Adler et al., 1999). It is impossible to decide whether the results were negative due to inactivity of the used drug or due to the poor performance of the AIMS when applied after 'life' examination in a long-term trial. This may illustrate that there is still a great need for well-constructed instruments to assess the severity of movement disorders on the long term. The SADIMoD may serve this purpose.

When we designed this study to investigate the characteristics of the SADIMoD, we tried to imitate the practical circumstances under which this instrument will be used. In our opinion, there is no use in knowing the test–retest reliability when it is measured with a time interval of only a few days or weeks or in knowing the inter-rater reliability when the individual raters have recently received an intensive training together. It is more informative to establish these characteristics under conditions that can easily be beaten in clinical trials. In the present study, the raters were not trained together, but separately by use of specific training material. Although it was possible to receive some additional training to refresh their memory, most raters were only trained once at the beginning of the study. As this was usually a short while before they scored the patients of their own centre for the first time, the time interval between training and assessing patients for establishing the inter-rater variability was quite long: usually several months (Loonen et al., 2000). In addition, two investigator teams were changed in between in which case the new members of that team were less experienced and were trained separately. Therefore, the results of the present study reflect the performance of the SADIMoD under very difficult circumstances. It can be expected that the training sessions that are organized during routine investigator's meetings in the everyday life situation of doing clinical research, will at least result in a similar reliability.

The inter-rater variability was established by calculating the Kendall's Coefficients of Concordance for the scorings of the six participating raters. Two different strategies were applied. First, Kendall's *W* was calculated for all separate subscale totals of the 31 patients. Secondly, Kendall's *W* was calculated for the 9 subscale totals of every individual patient. The SADIMoD was shown to perform very well. In all but two cases the scores correlated to a highly significant degree. For the sake of comparison: Gerlach et al. (1993) measured Kendall's *W* coefficients varying from 0.64 to 0.92 (median 0.84) in their study on the inter-rater variability of the SHRS. However, their seven raters were trained together and their study was completed within a 3-month period. Moreover, the SADIMoD procedure offers the opportunity to make video-recordings of patients at different moments and to have all of them scored in a very short period of time by a specific well-trained group of raters. This will probably decrease the inter-rater variability to a large extent.

A large data set with the SADIMoD scorings of a total of 112 patients was used to establish the construct validity. The high Cronbach's α coefficients of the separate subscales demonstrate that the subscale items were highly intercorrelated and that one item could be substituted for another. Even in the case of the Dystonia and Dyskinesia subscales these coefficients were far above the required 0.70. This is contrary to the expectation, that the homogeneity of the scale might be reduced due to occurrence of focal dystonias or to the separate occurrence of orofacial and peripheral dyskinesias. We did not consider the homogeneity of the Tremor subscale, as this scale quantifies three non-related types of tremors, i.e. rest tremor, postural tremor and intention tremor. The rather low coefficient for the Akathisia scale is probably largely due to the low number of items in this subscale. Moreover, the homogeneity may be lowered due to the occurrence of pseudoakathisia, that may be indistinguishable from dyskinesias or a tremor.

The convergent and divergent validity of the separate subscales largely met our expectations. Dyskinesias and dystonia on one hand and Parkinsonism, tremors, ataxia and akathisia on the other hand are likely to have the same dose–severity relationship. Peripheral dyskinesias may be indistinguishable from ataxia or (pseudo)akathisia, in which case they should both be scored according to the instructions of the SADIMoD. A rest tremor and postural tremor can be considered to be symptoms of Parkinsonism (M. P. Caligiuri, personal communication). The factor-analysis resulted in an excellent picture. The six factors, that were identified, corresponded very well with the intentions of the designers. This finding indicates that the SADIMoD is very well constructed.

Finally, the variability was established for individual patients who were assessed three times without any change of their clinical condition. This variability may be due to the testing procedure and to a day-to-day variability of the severity of movement disorders. The results show that the variability related to the test operations is low. Although the second and third scores were usually slightly higher than the first, none of the differences reached significance. The total variability appears to be largely due to intra-individual variability. The calculated standard errors can be used while planning future studies in order to perform a power analysis. From their (relatively) low correlation between the three ratings (Table 7) it can be concluded, that in this subset dystonia, postural tremor and intention tremor showed the largest intra-individual variability.

In conclusion, when investigated under realistic circumstances, the inter-rater reliability of the SADIMoD has been found to be satisfying. The instruction material, that was developed and used in this study, fully meets the requirements. The construct validity of the SADIMoD is excellent. This may make the SADIMoD currently the best available instrument to measure movement disorders in clinical trials in psychiatric patients. However, definite conclusions wait for its use by numerous independent research groups within the framework of conceptually and methodologically different protocols.

Acknowledgement

This project was funded by an unrestricted personal grant of Pfizer bv, Capelle aan den IJssel.

References

Adler LA, Edson R, Lavori Ph, Peselow E, Duncan E, Rosenthal M, Rotrosen J (1998). Long-term treatment effects of vitamin E for tardive dyskinesia. *Biological Psychiatry 43*, 868–872.

Barnes TRE (1989). A rating scale for drug-induced akathisia. *British Journal of Psychiatry 154*, 672–676.

Barnes TRE (1992). Neuromuscular effects of neuroleptics: akathisia. In: Kane JM, Lieberman JA (Eds.), *Adverse Effects of Psychotropic Drugs* (pp. 201–217). New York: Guildford Press.

Bartko JJ, Carpenter Jr. WT (1976). On the methods and theory of reliability. *Journal of Nervous and Mental Disease 163*, 307–317.

Bergen JA, Carter NB, Craig J, MacFarlane D, Smith EF, Beumont PJV (1988). AIMS ratings – repeatability. *British Journal of Psychiatry 152*, 670–673.

Bordeleau J-M, Albert J-M, Hillel J, Tetreault L (1967). Médication antiparkinsonienne et bilan extra-pyramidal: étude du trihexyphenidyl. *Canadian Psychiatric Association Journal 12*, 585–595.

Burke RE, Fahn S, Marsden CD, Bressman SB, Moskowitz C, Friedman J (1985). Validity and reliability of a rating scale for the primary torsion dystonias. *Neurology 35*, 73–77.

Caligiuri MP, Bracha HS, Lohr JB (1989). Asymmetry of neuroleptic-induced rigidity: development of quantitative methods and clinical correlates. *Psychiatry Research 30*, 275–284.

Chien C-P, DiMascio A, Cole JO (1974). Antiparkinsonian agents and depot phenothiazine. *American Journal of Psychiatry 131*, 86–90.

Chien C-P, Jung K, Ross-Townsend A, Stearns B (1977). The measurement of persistent dyskinesia by piezoelectric recording and clinical rating scales. *Psychopharmacology Bulletin 13*, 34–36.

Chouinard G, Ross-Chouinard A, Gauther S, Annable L, Mercier P (1984). An extrapyramidal rating scale for idiopathic and neuroleptic induced parkinsonism and dyskinesia. *14th CINP Congress*, Florence, Italy.

Crane GE, Ruiz P, Kernohan WJ, Wilson W, Royalty N (1969). Effects of drug withdrawal on tardive dyskinesia. *Activitas Nervosa Superior 11*, 30–35.

Cronbach LG (1951). Coefficient alpha and the internal structure of tests. *Psychometrika 16*, 197–334.

Cunningham Owens DG (1999). An overview of some standardised recording instruments. In: *A Guide to the Extrapyramidal Side-effects of Antipsychotic Drugs* (chapter 11). Cambridge: Cambridge University Press.

de Leon J, Simpson GM (1992). Assessment of neuroleptic-induced extrapyramidal symptoms. In: Kane JM, Lieberman JA (Eds.), *Adverse Effects of Psychotropic Drugs* (pp. 218–34). New York: Guildford Press.

Doller Wojcik J, Gelenberg AJ, LaBrie RA, Mieske M (1980). Prevalence of tardive dyskinesia in an outpatient population. *Comprehensive Psychiatry 21*, 370–380.

Doorschot CH, Doornbos M, Engberink MHA, Loonen AJM (1994). Een nieuw meetinstrument voor bewegingsstoornissen bij psychiatrische patiënten. *Spring Meeting of the Dutch Society of Psychiatry*, Lunteren.

Doorschot CH, Engberink M, Loonen AJM (1993). The stability of drug-induced movement disorders. *9th World Congress of Psychiatry*, Rio de Janeiro.

Edson R, Lavori Ph, Tracy K, Adler LA, Rotrosen J, Veterans Affairs Cooperative Study no. 394 Study Group (1997). Interrater reliability issues in multicenter trials. Part II: Statistical procedures used in department of veterans affairs cooperative study no. 394. *Psychopharmacology Bulletin 33*, 59–67.

Fahn S, Elton RL, and members of the UPDRS, Development Committee (1987). The unified Parkinson's disease rating scale. In: Fahn S, Marsden CD, Calne DB, Goldstein M (Eds.), *Recent Developments in Parkinson's Disease*, vol. 2 (pp. 153–164). Florham Park, NJ: Macmillan Health Care Information.

Fleischhacker WW, Bergmann KJ, Perovich R, Pestreich LK, Borenstein M, Lieberman JA, Kane JM (1989). The Hillside akathisia scale: a new rating instrument for neuroleptic-induced akathisia. *Psychopharmacology Bulletin 25*, 222–226.

Fleiss JL (1986). *The Design and Analysis of Clinical Experiments* (pp. 17–23). New York: John Wiley and Sons.

Gardos G, Cole JO, La Brie R (1977). The assessment of tardive dyskinesia. *Archives of General Psychiatry 34*, 1206–1212.

Gerlach J (1979). Tardive dyskinesia. *Danish Medical Bulletin 26*, 209–245.

Gerlach J, Korsgaard (1983). S. Classification of abnormal involuntary movements in psychiatric patients. *Neuropsychiatrica Clinica 2*, 201–208.

Gerlach J, Korsgaard S, Clemmesen P, Lund Lauersen A-M, 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*. Washington, DC: US Department of Health, Education, and Welfare.

Haase H-J (1977). *Therapie mit Psychopharmaka und anderen seelisches Befinden beeinflussenden Medikamenten*. Stuttgart: FK Schattauer Verlag.

Hershon HI, Kennedy PF, McGuire RJ (1972). Persistence of extra-pyramidal disorders and psychiatric relapse after withdrawal of long-term phenothiazine therapy. *British Journal of Psychiatry 120*, 41–50.

Hippius H, Logemann G (1970). Zur Wirkung von Dioxyphenylalanin (L-DOPA) auf extrapyramidalmotorische Hyperkinesen nach langfristiger neuroleptischer Therapie. *Arzneimittel-Forschung 20*, 894–896.

Hoehn MM, Yahr MD (1967). Parkinsonism: onset, progression, and mortality. *Neurology 17*, 427–442.

Jogems-Kosterman BJM, Loonen AJM, Van Hoof JJM, Zitman FG, Hulstijn W (1998). Antipsychotica, bewegingsstoornissen en fijne motoriek: nieuw instrumentarium voor het meten van bradykinesie en tremor. *Acta Neuropsychiatrica 10*, 93–99.

Kalachnik JE, Sprague RL. (1993). The dyskinesia identification system condensed user scale (DISCUS): reliability, validity, and a total cut-off for mentally ill and mentally retarded populations. *Journal of Clinical Psychology 49*, 177–189.

Kennedy PF, Hershon HI, McGuire RJ (1971). Extrapyramidal disorders after prolonged phenothiazine therapy. *British Journal of Psychiatry 118*, 509–518.

Klett CJ, Caffey E (1972). Evaluating the long-term need for antiparkinson drugs by chronic schizophrenics. *Archives of General Psychiatry 26*, 374–379.

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 Disorders 173*, 353–357.

Lingjærde O, Ahlfors UG, Bech P, Dencker SJ, Elgen K (1987). The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatrica Scandinavica 76* (Suppl. 334), 1–100.

Loonen AJM, Doorschot CH, Jogems-Kosterman BJM, Van Hoof JJM, Hulstijn W, Zitman FG (1997). Nieuw instrumentarium voor het meten van bijwerkingen. *COBO Bulletin 30*, 31–40.

Loonen AJM, Doorschot CH, Oostelbos MCJM, Sitsen JMA (1999). Lack of drug interactions between mirtazapine and risperidone in psychiatric patients: a pilot study. *European Neuropsychopharmacology 10*, 51–7.

Loonen AJM, Doorschot CH, Van Hemert DA, Oostelbos MCJM, Sijben AES, the MASEAS Team (2000). The schedule for the assessment of drug-induced movement disorders (SADIMoD): test–retest reliability and concurrent validity. *International Journal of Neuropsychopharmacology 2*, 285–296.

Loonen AJM, Engberink MHA, Doornbos M, Doorschot CH (1993) Scale for the Assessment of drug-induced movement disorders. *Meeting of the Dutch Society of Clinical Pharmacology and Biopharmacy*, Nijmegen.

Loonen AJM, Engberink MHA, Doornbos M, Doorschot CH (1994) Scale for the assessment of drug-induced movement disorders (SADIMoD) – optimalization of the video-registration procedure. *XIXth CINP Congress*, Washington DC, USA.

Loonen AJM, Verwey HA, Roels PR, van Bavel LP, Doorschot CH (1992). Is diltiazem effective in treating the symptoms of (tardive) dyskinesia in chronic psychiatric inpatients? A negative, double-blind, placebo-controlled trial. *Journal of Clinical Psychopharmacology 12*, 39–42.

Loonen AJM, Zwanikken GJ (1987). *Farmaca, Psychofarmaca, Klinisch Onderzoek*. 's-Hertogenbosch: Wulco.

Lord FM, Novick MR (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Marder SR, van Putten T, Mintz J, McKenzie J, Lebell M, Faltico G, May PRA (1984). Costs and benefits of two doses of fluphenazine. *Archives of General Psychiatry 41*, 1025–1029.

Mindham RHS (1976). Assessment of drug-induced extrapyramidal reactions and of drugs given for their control. *British Journal of Clinical Pharmacology* (Suppl.), 395–400.

Nunnally Jr. JC (1970). *Introduction to Psychological Measurement*. New York: McGraw-Hill.

Rabey JM, Bass H, Bonuccelli U, et al. (1997). Evaluation of the short Parkinson's evaluation scale: a new friendly scale for the evaluation of Parkinson's disease in clinical drug trials. *Clinical Neuropharmacology 20*, 322–337.

Reda FA, Escobar JI, Scanlan JM (1975). Lithium carbonate in the treatment of tardive dyskinesia. *American Journal of Psychiatry 132*, 560–562.

Sachdev P (1994). A rating scale for acute drug-induced akathisia: development, reliability, and validity. *Biological Psychiatry 35*, 263–271.

Simpson GM, Amuso D, Blair JH, Farkas T (1964). Aspects of phenothiazine-produced extrapyramidal symptoms. *Archives of General Psychiatry 10*, 199–208.

Simpson GM, Angus JWS (1970). A rating scale for extrapyramidal side effects. *Acta Psychiatrica Scandinavica 45* (Suppl. 212), 11–19.

Simpson GM, Lee JH, Zoubok B, Gardos G (1979). A rating scale for tardive dyskinesia. *Psychopharmacology 64*, 171–179.

Simpson GM, Lindenmayer JP (1997). Extrapyramidal symptoms in patients treated with risperidone. *Journal of Clinical Psychopharmacology 17*, 194–201.

Sinnaeve AF (1989). Het meten van tremor. *Techniek Gezondheidszorg 5*, 10–12.

Smith JM, Kucharski LT, Eblen C, Knutsen E, Linn C (1979b). An assessment of tardive dyskinesia in schizophrenic outpatients. *Psychopharmacology 64*, 99–104.

Smith JM, Kucharski LT, Oswald WT, Waterman LJ (1979a). A systematic investigation of tardive dyskinesia in inpatients. *American Journal of Psychiatry 136*, 918–922.

Smith RC, Allen R, Gordon J, Wolff J (1983). A rating scale for tardive dyskinesia and parkinsonian symptoms. *Psychopharmacology Bulletin 19*, 266–276.

Smith RC, Tamminga CA, Haraszti J, Pandey GN, Davis JM (1977). Effects of dopamine agonists in tardive dyskinesia. *American Journal of Psychiatry 134*, 763–768.

Sweet RA, DeSensi EG, Zubenko GS (1993). Reliability and applicability of movement disorder rating scales in the elderly. *Journal of Neuropsychiatry and Clinical Neurosciences 5*, 56–60.

Teräväinen H, Tsui JKC, Mak E, Calne DB (1989). Optimal indices for testing parkinsonian rigidity. *Canadian Journal of Neurological Sciences 16*, 180–183.

Tracy K, Adler LA, Rotrosen J, Edson R, Lavori P, and the Veterans Affairs Cooperative Study no. 394 Study Group (1997). Interrater reliability issues in multicenter trials. Part 1: Theoretical concepts and operational procedures used in department of veterans affairs cooperative study no. 394. *Psychopharmacology Bulletin 33*, 53–57.

Van Eindhoven HPJ, Loonen AJM, Doorschot CH, Zitman FG (1995). Variabiliteit van bewegingsstoornissen bij chronisch psychiatrische patiënten. Onderzoeksverslag wetenschappelijke stage. Vught: Psychiatric Hospital Reinier van Arkel.

Webster DD (1968). Critical analysis of the disability in Parkinson's disease. *Modern Treatment 5*, 257–282.