Interrater Reliability of Extrapyramidal Signs in a Group Assessed for Dementia

Marcus Richards, PhD; Karen Marder, MD, MPH; Karen Bell, MD;
George Dooneief, MD; Richard Mayeux, MD; Yaakov Stern, PhD

• Extrapyramidal signs were rated by three neurologists in 20 patients who had either been diagnosed as having probable Alzheimer's disease or who were being evaluated for dementia. In general, good interrater reliability was found for the presence or absence of extrapyramidal signs, although agreement over the presence of some signs was reduced when distinctions between normality and slight departures from normality were required.

(*Arch Neurol.* 1991;48:1147-1149)

S everal studies have shown that mild extrapyramidal signs (EPSs) are frequently observed in Alzheimer's disease (AD).[1-5] Their presence in AD is significant for at least two reasons. First, it has been shown that EPS in AD, particularly rigidity, may correspond with the presence of neuropathologic parkinsonian features (eg, Lewy body formation, neuronal loss and gliosis of the substantia nigra, and locus ceruleus).[6] Extrapyramidal sign presence may therefore have etiologic and pathophysiologic implications for AD. Second, there is evidence that the presence of EPS may signify an aggressive form of AD[7,8] and may thus be an important predictor of disease severity. However, the validity of EPS as clinical correlates and as disease predictors depends critically on the reliability of their quantitative assessment. Reliability is defined as the degree of agreement either among the same rater at different times (intrarater reliability) or among different raters at the same time (interrater reliability). Assessment of intrarater reliability of EPS is rendered difficult by the fact that, in our opinion, the presentation of certain mild EPS within patients is not constant over time. It is therefore essential to assess interrater reliability of EPS for small comparative studies where replicability is important and for larger cohort and population-based studies where multiple raters are

often employed. Interrater reliability for neurologic soft signs in children, including tone[9] and motor slowness,[10] has been established. However, to date, no attempts have been made to assess interrater reliability of EPS in a dementia cohort, although there are several studies where such information would be crucial. This study undertook such an assessment with patients who were either undergoing evaluation or receiving ongoing care for dementia.

PATIENTS AND METHODS

A convenient sample of 20 patients were assessed at a neurologic outpatient facility designed to assess memory disorders. Sixteen of these patients were diagnosed as having probable AD, one was diagnosed with vascular-related dementia, and three had cognitive changes of unknown cause at the time of assessment. Ages ranged from 55 to 87 years, with a mean of 72.2 years. Ratings were performed by three of us (K.M., K.B., and G.D.), each with approximately 4 years of postresidency experience in behavioral

Table 1.—Extrapyramidal Signs (EPS) Items and Scale Construction Used in the Present Study

EPS Items	Scale
Speech	
Normal	0
Slight loss of expression, diction and/or volume	1
Monotone, slurred but understandable; moderately impaired	2
Marked impairment, difficult to understand	3
Unintelligible	4
Tremor at rest	
Absent	0
Slight and infrequently present	1
Mild in amplitude and present most of the time	2
Moderate in amplitude and present most of the time	3
Marked in amplitude and present most of the time	4
Facial expression	
Normal	0
Slight hypomimia, could be normal "poker face"	1
Slight but definitely abnormal diminution of facial expression	2
Moderate hypomimia; lips parted some of the time	3
Masked or fixed face with severe or complete loss of facial expression; lips parted ¼ inch or more	4
Rigidity (neck and each limb) *	
Absent	0
Slight or detectable only when activated by mirror	1
Mild to moderate	2
Marked, but full range of motion easily achieved	3
Severe, range of motion achieved with difficulty	4
Posture	
Normal erect	0
Not quite erect, slightly stooped; could be normal for older person	1
Moderately stooped posture, definitely abnormal; can be leaning slightly to one side	2
Severely stooped posture with kyphosis; can be leaning moderately to one side	3
Marked flexion with extreme abnormality of posture	4
Gait	
Normal	0
Walks slowly, may shuffle with short steps, but no festination or propulsion	1
Walks with difficulty, but requires little or no assistance; may have some festination, short steps or propulsion	2
Severe disturbance of gait, requiring assistance	3
Cannot walk at all, even with assistance	4
Body bradykinesia and hypokinesia†	
None	0
Minimal slowness, giving movement deliberate character; could be normal for some persons. Possibly reduced amplitude	1
Mild degree of slowness and poverty of movement that is definitely abnormal Alternatively, some reduced amplitude	2
Moderate slowness, poverty, or small amplitude of movement	3
Marked slowness, poverty, or small amplitude of movement	4

* Judged on passive movement of major joint with patient relaxed in sitting position. Cogwheeling to be ignored.

†Combining slowness, hesitancy, decreased arm swing, small amplitude, and poverty of movement in general.

| | Speech | Facial Immobility | Tremor | Rigidity | | | | | | Posture | Bradykinesia/ Hypokinesia | Gait |
				Neck	R Arm	L Arm	R Leg	L Leg	Overall			
N	19	20	20	17	19	19	19	19	19	20	20	20
No. with sign present	5	7	0	4	9	7	6	4	11	4	5	3
Frequency, %	26	35	0	21	45	35	30	20	55	20	25	15
% Observed Agreement	90	60	95	82	68	68	74	90	63	75	85	85
% Chance agreement	37	32	95	53	27	33	42	44	24	43	41	58
Kappa	.83	.41	0	.63	.57	.53	.55	.81	.52	.56	.75	.64
χ^2	41.2	14.1	...	20.5	20.5	16.0	17.1	38.0	16.8	20.8	35.9	28.4

*N = 20.

| | Speech | Facial Immobility | Tremor | Rigidity | | | | | | Posture | Bradykinesia/ Hypokinesia | Gait |
				Neck	R Arm	L Arm	R Leg	L Leg	Overall			
N	19	20	20	17	19	19	19	19	19	20	20	20
No. with sign present	2	3	0	2	3	2	3	3	3	1	3	2
Frequency, %	10	16	0	10	16	10	16	16	16	5	16	10
% Observed Agreement	90	90	100	100	100	95	100	95	100	95	85	100
% Chance agreement	72	77	100	69	60	64	60	64	60	81	69	73
Kappa	.63	.57	0	1	1	.85	1	.78	1	.73	.52	1
χ^2	22.4	24.4	...	51.0	57.0	42.8	57.0	42.8	57.0	35.0	16.5	60.0

*N = 20.

neurology. The EPSs were assessed using selected items from the Columbia University Parkinson's Disease Rating Scale.[11] This scale enables separate ratings of voice, facial immobility, resting tremor, rigidity (neck and each limb), bradykinesia/hypokinesia, posture, and gait (Table 1) in a semistructured fashion.

These items were selected on the basis of an earlier characterization of EPS in dementia.[12] Sign severity was rated as either absent (0), slight (1), mild-moderate (2), marked (3), or severe (4). The presence or absence of chorea, myoclonus, and other dyskinesias was also determined. Rigidity assessments were performed by each rater in immediate succession, with rater order randomized. All other signs were rated by observation. To determine whether signs were idiopathic or drug-induced, all raters were acquainted with the medication history of each patient prior to assessment. In all other respects, ratings were made independently and raters remained unaware of each other's assessments until the completion of the study.

DATA ANALYSIS

The frequency of each sign (including an overall present/absent rigidity score) was calculated on the basis of assessments by the initial rater for each patient. Because the frequency of EPS with a severity rating greater than 1 (slight) was small, the initial analysis focused on the presence or absence of EPS only. To assess interrater agreement, four statistics were calculated. First, percent agreement between raters for each sign was expressed as the ratio between the num-

ber of patients where all three raters agreed on presence or absence of that sign and the total number of patients. However, the magnitude of this simple statistic may be inflated by chance agreement, especially when signs occur either very rarely or very frequently in a sample.[13] Percent agreement by chance alone, derived from the expected frequencies for a χ^2 test, are therefore shown for each sign. Third, kappa coefficients were calculated for each sign. The kappa statistic adjusts the simple agreement ratio for chance agreement by measuring the ratio of the observed agreement beyond chance to the potential agreement beyond chance. Thus a kappa of 1.0 represents complete agreement beyond chance, a kappa of 0 indicates agreement at the chance level and a kappa of -1.0 represents complete disagreement between raters. Fleiss[14] has suggested the following guidelines for interpreting kappas: coefficients with values greater then .75 signify excellent agreement, values between .40 and .75 represent fair to good agreement, and values below .40 or so indicate poor agreement beyond chance. Finally, to assess the significance of agreement beyond chance, χ^2 values were calculated for each sign, based on the ratio between kappa magnitude and rater variance and thus taking sample size into account.[15]

RESULTS

All three raters agreed on the presence of myoclonus in one patient and on the absence of chorea and other dyskinesias in all patients. Based on assessments by the initial rater for each pa-

tient, 13 patients (65%) had at least one EPS. Of these, signs in 10 patients were idiopathic, signs in two patients were judged to be induced by current neuroleptic medication, and signs in one patient were considered as possible residual effects of past medication. The frequency of each sign, along with percent chance agreement, percent observed agreement, and kappa coefficients for these signs, are presented in Table 2. Raters were unable to induce one patient to talk during the examination. In addition, individual raters were unable to relax three patients sufficiently to assess neck rigidity; for the same reason, one rater was unable to rate limb rigidity in one patient. In each case, the above signs for these patients were dropped from the analysis.

It can be seen that EPSs were frequent in the present group of patients. In particular, over half the sample showed rigidity in at least one area of assessment. Bradykinesia/hypokinesia, abnormal speech, and facial immobility were also common. Gait abnormalities were less frequently observed, however, and there were no instances of resting tremor. Table 2 indicates high agreement above chance between raters on the presence or absence of EPSs for speech, tremor, bradykinesia/hypokinesia, and gait, with kappas in the "good" to "excellent" range.[14] Lower

agreement above chance was found for facial immobility, posture, and overall rigidity, with considerable variability in agreement over rigidity across the five individual areas assessed. Kappas for these signs ranged from "poor" (facial immobility), to "fair" (overall rigidity and posture).[14] Agreement was no better than chance, however, for resting tremor, where a single instance of positivity was reported by one rater. Where target events are very rare, the probability of chance agreement is high; in this case, percent chance agreement was equal to percent observed agreement, yielding a kappa of 0 for this sign. All χ^2 values were significant at the .001 level or less (with the exception of that for tremor), indicating statistically significant levels of agreement beyond chance for most signs with this sample.

One possible reason for the relatively modest agreement over some of these signs is the ambiguous status of the "slight" (severity rating, 1) level of severity. For example, the guidelines for assigning facial and postural signs to the slight level are worded to include features that may be normal for older people. To create a more conservative absent/present dichotomy therefore, slight EPS scores were reassigned to the absent category and EPSs were considered as present only if their severity was at least mild to moderate (severity rating, 2). Their frequencies, along with corresponding interrater reliability estimates, are shown in Table 3.

The frequency of EPS of at least a mild to moderate level of severity was significantly lower than the frequency of signs where presence included the slight level. Kappas for speech and bradykinesia/hypokinesia were lower than corresponding coefficients in Table 2 (although still in the "fair to good" range[14]); this is because the frequency of positivity is lower and the probability of chance agreement is higher for signs determined by this new dichotomy. However, kappa values increased for all other signs except tremor (where all raters were agreed on sign absence for all patients and, again, percent chance agreement was equal to percent observed agreement). Notably, there was total agreement beyond chance for overall rigidity and gait. Once again, χ^2 values were significant at the .001 level or less for all signs except tremor.

COMMENT

Even though the sample size was small, our study demonstrates an acceptable degree of agreement between raters over the presence or absence of mild EPS in this group. This finding offers evidence that ratings of EPS are

capable of standardization and are reliable and reproducible. This has important bearings on the validity of EPS as manifestations of pathophysiologic changes in the central nervous system, as correlates of functional and neuropsychologic changes in dementia, and as predictors of disease severity. From a methodologic point of view, these results also support the employment of multiple raters in epidemiologic studies of AD and other dementias, and suggest that individual EPS ratings in small-scale studies are capable of generalization. Moreover, in light of the fact that clinical teaching depends largely on demonstration and observation in a group setting, these results reinforce the view that a consensus over the presence of mild EPS can be reached in a clinical setting.

This study does, however, highlight two problems in the assessment of reliability of EPS detection. First, since the probability of chance agreement is very high for signs that occur rarely in AD (such as resting tremor), estimates of reliability that correct for chance agreement are likely to be of low magnitude for these signs. This does not mean that reliability in the assessment of these signs is unattainable, merely that adequate assessment of reliability cannot be made under such circumstances because it is difficult to determine the ability of raters to distinguish targets.

Second, and in line with investigations of soft signs in children,[16] interrater agreement was shown to be disrupted by distinctions between normality and slight deviations from normality with some signs. A more reliable measure of EPS presence may therefore be provided by considering only EPS with a severity greater than slight (as in analysis two above). In line with this suggestion, Thompson and Walter[17] have demonstrated that when sensitivity (ie, the proportion of targets correctly classified as present) and specificity (ie, the proportion of targets correctly classified as absent) are held constant, the magnitude of kappa for any given target declines as the prevalence of that target declines beyond an optimum level. In our study, the frequency of signs with a severity greater than slight was substantially lower than the frequency of signs with a severity greater than zero, yet kappa values for the former were generally larger than for the latter. This suggests that increasing the threshold for EPS presence from slight (severity rating, 1) to mild (severity rating, 2) led to an improvement in the sensitivity and specificity of these measures.

At the same time, the potential diffi-

culties in making distinctions between normality and slight deviations from normality warrant continuing attention. It is therefore important when assessing mild EPSs in older patients to clarify the range of behaviors that may be considered normal for this age group. Not only is this important for reliability per se, but it is also only against such a baseline that the clinical significance of mild positive EPSs can be fully determined.

References

1. Pearce J. The extrapyramidal disorder of Alzheimer's disease. Eur Neurol. 1974;12:94-103.
2. Mayeux R, Stern Y, Spanton S. Heterogeneity in dementia of the Alzheimer type: evidence of subgroups. Neurology. 1985;35:453-461.
3. Funkenstein HH, Albert M, Cook N, West C, Scherr P, Evans B. Relationships of neurological examination findings to diagnosis of SDAT in a defined community population. Neurology. 1986;36(suppl 1):267.
4. Huff HJ, Boller F, Lucchelli F, Querriera R, Beyer J, Belle S. The neurologic examination in patients with probable Alzheimer's disease. Arch Neurol. 1987;44:929-932.
5. Girling DM, Berrios GE. Extrapyramidal signs, primitive reflexes and frontal lobe function in senile dementia of the Alzheimer type. Br J Psychiatry. 1990;157:888-893.
6. Ditter SM, Mirra SS. Neuropathologic and clinical features of Parkinson's disease in Alzheimer's disease patients. Neurology. 1987;37:754-760.
7. Chui HC, Teng EL, Henderson VW, Moy AC. Clinical subtypes of dementia of the Alzheimer type. Neurology. 1985;35:1544-1550.
8. Stern Y, Mayeux R, Sano M, Hauser WA, Bush T. Predictors of disease course in patients with probable Alzheimer's disease. Neurology. 1987;37:1649-1653.
9. Rutter M, Graham P, Yule W. A neuropsychiatric study in childhood. In: Clinics in Developmental Medicine. Philadelphia, Pa: JB Lippincott; 1970.
10. Stokman CJ, Shafer SQ, Shaffer D, Ng S, O'Connor PA, Wolff RR. Assessment of neurological 'soft signs' in adolescents: reliability studies. Dev Med Child Neurol. 1986;28:428-439.
11. Stern MB. The clinical characteristics of Parkinson's disease and Parkinsonian syndromes: diagnosis and assessment. In: Stern MB, Hurtig HI, eds. The Comprehensive Management of Parkinson's Disease. New York, NY: PMA Publishing Corp; 1988:3-50.
12. Pearce J. The extrapyramidal disorder of Alzheimer's disease. Eur Neurol. 1974;12:94-103.
13. Longstreth W, Koepsell TD, Van Belle G. Clinical neuroepidemiology, I: diagnosis. Arch Neurol. 1987;44:1091-1099.
14. Fleiss JL. Statistical Methods for Rates and Proportions. 2nd ed. New York, NY: John Wiley & Sons Inc; 1981:218.
15. Ross DC. Testing patterned hypotheses in multi-way contingency tables using weighted kappa and weighted chi square. Educ Psychol Measur. 1977;37:291-307.
16. Yule W, Taylor E. Classification of soft signs. In: Tupper DE, ed. Soft Neurological Signs. New York, NY: Grune & Stratton; 1987:25.
17. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. J Clin Epidemiol. 1988;41:949-958.