Multiple Sclerosis 2003; **9:** 154–159
www.multiplesclerosisjournal.com

Can the Expanded Disability Status Scale be assessed by telephone?

J Lechner-Scott[1,2], L Kappos[*,1], M Hofman[1], CH Polman[3], H Ronner[3], X Montalban[4], M Tintore[4], M Frontoni[5], C Buttinelli[5], MP Amato[6], ML Bartolozzi[6], M Versavel[7], F Dahlke[7], J-F Kapp[7] and R Gibberd[8]*

[1]Department of Neurology, Kantonspital Basel, Switzerland; [2]Department of Medicine, John Hunter Hospital, Newcastle, Australia; [3]Department of Neurology, University Hospital, Amsterdam, Netherlands; [4]Department of Neurology, Ciudad Sanitaria Vall d'Hebron, Barcelona, Spain; [5]Clinica Neurologica de l'Università Rome, Rome, Italy; [6]Department of Neurological and Psychiatric Sciences, University of Florence, Italy; [7]Klinische Entwicklung Schering, Germany; [8]Department of Statistics, University of Newcastle, Newcastle, Australia

Information from patients who are unable to continue their visits to a study centre may be of major importance for the interpretation of results in multiple sclerosis (MS) clinical trials. To validate a questionnaire based on the Expanded Disability Status Scale (EDSS), patients in five different European centres were assessed independently by pairs of trained EDSS raters, first by telephone interview and a few days later by standardized neurological examination. Seventy women and 40 men with an average age of 43.7 years (range 19–74 years) were included in the study. Mean EDSS score at the last visit was 4.5 (0–9). EDSS assessment by telephone was highly correlated with the EDSS determined by physical examination (Pearson's correlation coefficient = 0.95). An intraclass correlation coefficient (ICC) of 94.8% was found for the total sample; 77.6% and 86%, respectively, for patients with EDSS < 4.5 (n = 46) and > 4.5 (n = 64). Kappa values for full agreement were 0.48; for variation by ±0.5 steps and ±1.0 steps, 0.79 and 0.90, respectively. Best agreement could be found in higher EDSS scores, where assessment by telephone interview might be needed most. The telephone questionnaire is a valid tool to assess EDSS score in cases where the patient is unable to continue visiting a study centre or in long-term follow-up of trial participants.
Multiple Sclerosis *(2003)* **9,** 154–159

Key words: *assessment; clinical scale; multiple sclerosis*

Introduction

In multiple sclerosis (MS) clinical trials both the statistical power to detect an existing difference between treatments and the applicability of the results to the general population may be severely compromised by missing data and incomplete follow-up of patients. Especially in higher grades of disability or in studies with long duration, the inability of patients to appear for regular examinations in the study centre may be an important factor. Sometimes emergency visits for relapses are not possible because the patient is unable to come to the study centre or the study centre is unable to schedule the visit in the appropriate timeframe. To collect information about the degree of deterioration, a telephone-based assessment may be a better option than assessing the patient when the relapse has resolved.

Therefore, a structured questionnaire was developed, which enables the study centre to assess disease progression by telephone interview in the absence of a physical examination of the patient. The questionnaire is based on the Expanded Disability Status Scale (EDSS),[1] a measure accepted as gold standard[2] (although much criticized)[3] and used in most clinical trials as primary clinical outcome. In this multicentre study, EDSS assessment by telephone is compared with EDSS assessment based on a standardized physical neurological examination including assessment of walking distance.[4]

**Correspondence: L Kappos, Outpatient Clinic Neurology-Neurosurgery, University Hospitals Basel, Petersgraben 4, CH-4031 Basel, Switzerland.*
E-mail: lkappos@uhbs.ch
Received 24 January 2001; revised 17 July 2002; accepted 18 July 2002

Material and methods

Study design

Patients with clinically definite MS, with at least one well-documented physical neurological examination at the study centre, including EDSS assessment, in the previous 12 months, were eligible for this study if they had not experienced a relapse or change in ongoing MS medication within the last month. After informed consent, half of the patients were interviewed using the EDSS telephone questionnaire by one of two trained EDSS physicians of the respective centre, and the other half was interviewed by the second examiner. Both EDSS physicians were aware of the previous (baseline) EDSS score. The interview was performed within a maximum of five days before a scheduled physical examination at the study centre. At this scheduled examination, the other EDSS physician performed a standardized neurological

10.1191/1352458503ms884oa

examination, including walking with the patient up to 500 m, and assessed the functional systems and the EDSS without knowledge of the telephone interview results. All EDSS raters in this study had undergone a standardized audiovisual training for EDSS assessment.[4]

Description of the EDSS by telephone questionnaire

(The full questionaire and guidelines for its administration are available through the corresponding author.)

The classical assessment of the EDSS is based on a neurological examination that provides the basis for the assessment of several functional systems that contribute according to predefined algorithms to the total EDSS score. For an EDSS score of 4.5 and above, walking ability and the need for assistance play a dominant role in determining the final score and the contribution of functional systems diminishes. Therefore, in the structured interview most questions (3–10a) are centred round these issues: walking distance, need of assistance and transfer. If the EDSS is sufficiently defined by the degree of dependence, assessment of single functional systems is not required.

The first question determines if the patient is interviewed himself or the information is given by a caregiver. The second question assesses any subjective change to the last examination at the centre.

The functional systems are assessed in question 10b. To illustrate the use of the structured interview a few items were picked for more detailed description:

Brainstem functional system (FS) comprises a number of tests in the physical examination. In the questionnaire the patient is asked if they have noticed any double vision, dysarthria, loss of sensation in the face, facial asymmetry or frequent swallowing problems. If an impairment exists further questions have to determine its degree. For example, if the patient has not lost but has only impaired sensation of one side of the face, this would give them a score of 2 out of 4. If the patient suffers, for example, from incomprehensible dysarthria, this would result in a score of 4.

Rating of the Pyramidal FS is based on which and how many limbs are involved in both the questionnaire and the physical examination.

Cerebellar FS is assessed by two questions in the interview related to limb ataxia and balance on ordinary walking.

While the physical examination assesses four qualities of sensation, the interview asks about impaired perception of touch or pain being the most apparent sensory deficit.

In the assessment of the Bowel/bladder and Cerebral FS, both the neurological examination and interview usually rely on patient information.

The final EDSS score is calculated in the same way as the EDSS score in the physical examination.

Statistical evaluation

The aim of this study was to determine the values of the telephone EDSS in relation to the subsequent physical examination EDSS. In particular, it was desired to determine the bias and the precision of the telephone interview given the physical examination score. This is commonly done by presenting the scatter plot of two scores,[5] but the mean and standard deviation of the differences in the two scores are also graphically presented: namely the telephone EDSS score minus the physical EDSS score, plotted against the physical EDSS score, which is assumed to be the gold standard. This allows one to determine the bias and precision of the telephone interview.

Because the EDSS is an ordinal variable, the kappa value was used to measure agreement as well as the intraclass correlation coefficient (ICC), which can be used for measurement of ordinal and nominal data. The kappa coefficient provides a measure of the degree of interobserver agreement for pairs of observers assigning individuals subjectively to one of a range of categories correcting for the extent of agreement expected by chance. It is the most widely used index for agreement.[6] The kappa value is interpreted as poor for <0.0, slight for 0.0–0.2, fair for 0.21–0.4, moderate for 0.41–0.6, substantial for 0.61–0.8 and almost perfect for 0.81–1.0.[6] To estimate the agreement of the single FS, the comparison was restricted to patients with EDSS scores of <4.5, because above this value FS are not necessary to determine the EDSS score and data were incomplete. Therefore, the ICC and kappa values for FS were calculated for 46 cases with an EDSS score <4.5.

The value 1 in most of the original FS according to Kurtzke corresponds to just signs in the physical examination of which the patient is not aware. It is not possible to assess physical signs in a telephone interview that do not result in symptoms. The telephone interview, therefore, does not differentiate between FS scores of 0 and 1.

Results

The patient group consisted of 70 women and 40 men, average age was 43.7 years (range 19–74 years). The mean duration of disease was 12 years (range 1–25 years). The patients presented with a wide range of baseline EDSS scores (Figure 1) at the last visit (mean 4.5), which was on average 3.36 months previous (range 0.5–6 months).

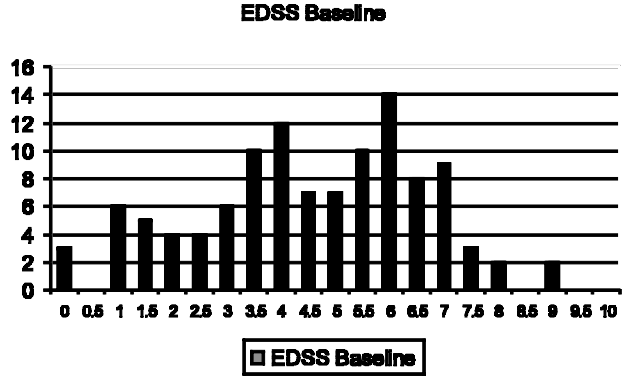

Figure 1 *Distribution of patients according to last prestudy (baseline) EDSS scores ($n=110$)*

Forty-two patients had a relapsing remittent course, 13 a progressive relapsing, 47 a secondary progressive and eight a primary progressive course.[7]

EDSS assessment by telephone correlated closely with the EDSS determined by physical examination (correlation coefficient for EDSS = 0.95).

The same close correlation was observed if the five participating centres were evaluated separately.

Four patients could not be interviewed directly. The caregiver (relative or professional) answered the questions instead. Their EDSS score varied widely (4.0, 5.5, 6.5 and 9.0). Only two of the four patients had some cognitive impairment. The statistical significance of these results did not change if these four patients were excluded from analysis (data not shown).

The mean of the difference in the two EDSS scores was unrelated to the mean obtained from the physical EDSS score, except for when the physical score was 9. In this case, the telephone EDSS score was 0.5 less than the physical score.

Figure 2 shows that the standard deviation of the differences depended on the value of the physical score. The larger the physical score the more precise was the telephone value. For scores of 2.5 and above, 50% or more of the cases were within 0.5 of each other, while for scores between 1 and 2, there was greater spread about the physical scores. To some extent, this can be explained by the values obtained by the physical EDSS scores: there were no values of 1.0 and 1.5 in the telephone interview, while there were for the physical EDSS scores. The telephone method resulted in these values being coded as 0.0 or 2.0 or more. It was this aspect that resulted in a lower correlation coefficient in the lower EDSS scores, even though the results were unbiased (the telephone EDSS differed as often positively as negatively from the physical EDSS). Carrying out an analysis of variance (ANOVA) on the data indicated that there was no statistically significant difference in the means ($P > 0.9$), but there was a statistically significant difference in the variances (Bartlett test for equal variances, $P < 0.0001$).

The ICC was 0.78 for cases with a telephone EDSS score of 4 or less, but was 0.87 for those above 4.0. This is consistent with Figure 2, showing that the agreement is better for the higher scores. The overall ICC was 0.95 for all cases.

Looking at the impact of mental impairment, there were 61 cases with Mental FS equal to zero and 49 not equal to zero. Comparing the two groups, a statistically significant difference in agreement between the EDSS scores could not be detected (ICC = 0.95 and 0.93, respectively).

The kappa value for full agreement was 0.48, for an EDSS score between ± 0.5 the value was 0.79 and ± 1, 0.90, which corresponds to almost perfect agreement. Best agreement was found for EDSS scores of 6–6.5 (100% agreement in 20 patients).

To achieve consistent results the neurologists were trained to walk a measured distance with the patient up to 500 m with as little assistance as possible. The assessment of ambulation correlated only modestly (correlation coefficient = 0.62; Figure 3). Nevertheless, the walking distance had a high predictive value for the outcome of the EDSS, as expected. The agreement for the single FS varied substantially (Table 1). Poor correlation was found for Mental (ICC = 0.30) and Cerebellar FS (ICC = 0.32). Best agreement was found for Brainstem and Pyramidal FS (ICC = 0.78 and 0.66, respectively).

Discussion

The study population of 110 patients is consistent with the demographic characteristics of Weinshenker's geographically based survey.[8] He described 1099 patients of which 722 were female. In the current population, 70 of 110 were female. Average duration of disease was 12 years in both populations. The average age at disease onset was

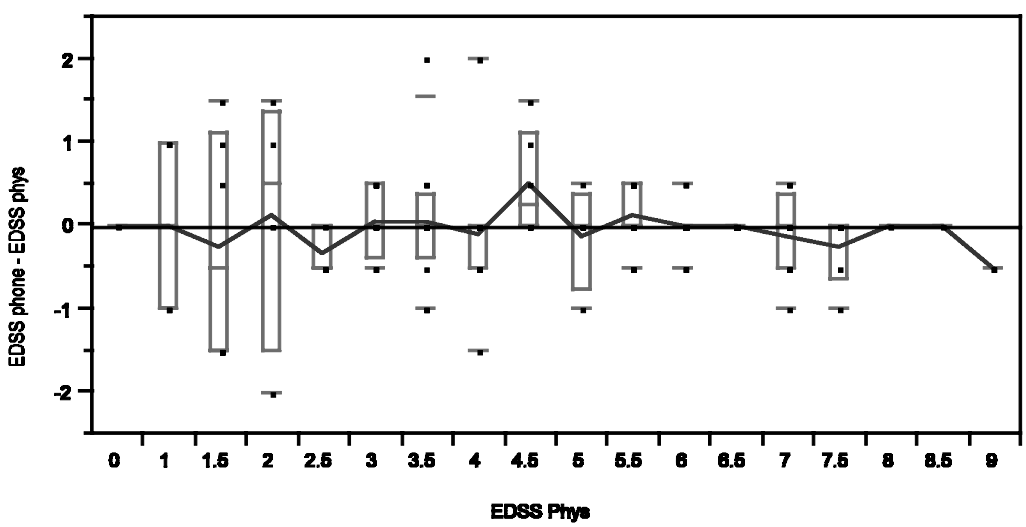

Figure 2 *Relationship between the difference in the EDSS by telephone and the physical EDSS, plotted against the physical EDSS. The line represents the mean difference and the boxes contain 50% of the observations. The agreement between the telephone and physical scores is best for higher values of the physical score, with more variation for scores 1, 1.5 and 2.*

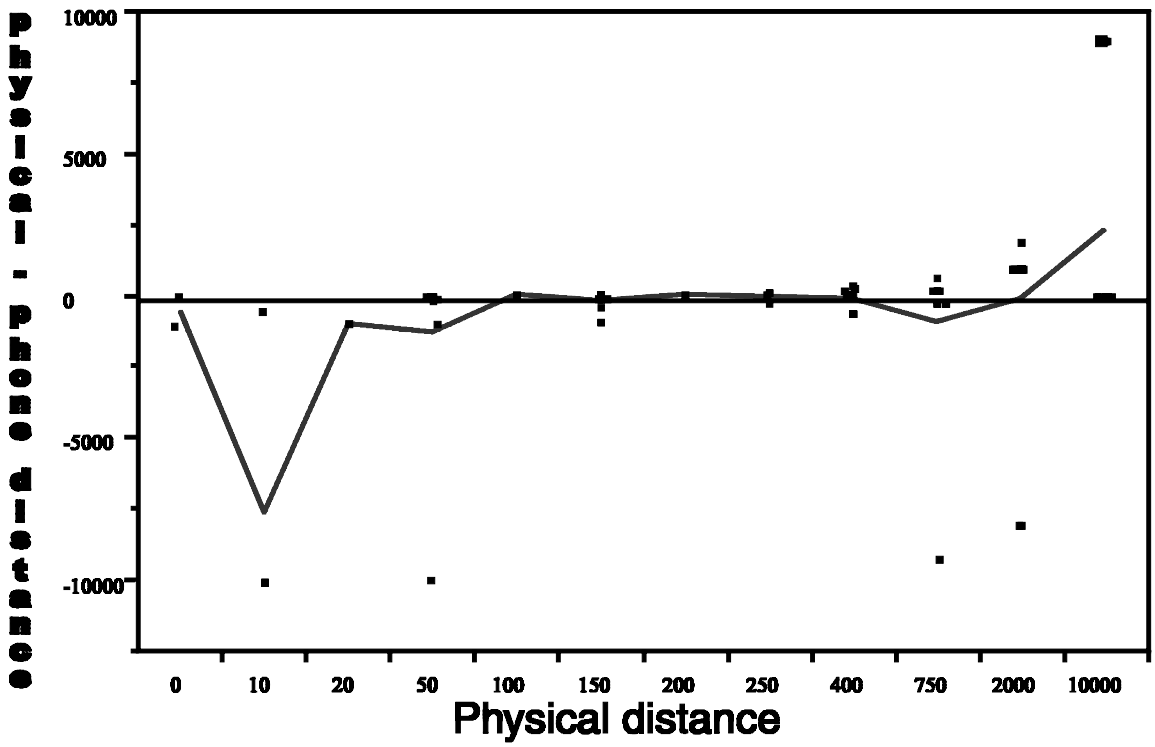

Figure 3 *Relationship between the difference in the walking distance from the physical examination and the telephone interview plotted against the distance from the physical examination. The line represents the mean difference.*

Table 1 Agreement for single FS between EDSS by telephone and by physical examination ($n = 46$, only patients with EDSS score < 4.5)

FS	*Kappa*	*Standard error (SE)*	*ICC (%)*	*SE*
Visual	0.22	0.09	0.38	0.05
Brainstem	0.59	0.12	0.78	0.07
Pyramidal	0.54	0.11	0.66	0.09
Cerebellar	0.40	0.11	0.32	0.07
Sensory	0.30	0.10	0.38	0.06
Bowel/bladder	0.42	0.11	0.43	0.08
Mental	0.38	0.13	0.30	0.06

31 years. Also, the dual distribution of EDSS has been described before.[3] EDSS scores between 3–4 and 6–6.5 are more frequent than EDSS scores of 5. It can be confidently assumed that the current study population represented the normal variation of disease.

In order that the results can be compared with the results of other authors, Pearson's correlation coefficient, ICC and the kappa coefficient were evaluated. These measures are used for inter-rater agreement, but in this study they cover both the agreement between two methods of measurement and between pairs of raters. In general, good inter-rater agreement is considered more difficult to obtain than intrarater agreement. Although the raters in this study had undergone standardized training,[4] which reduces interrater variability, it can not exactly be determined to what extent interrater variability reduced the agreement between the two measures.

Two other studies have addressed the feasibility of self-assessment of neurological impairment in MS and its correlation with physical EDSS. Goodin[9] reported 30 patients examined by himself. He describes a high Pearson correlation coefficient (0.92) between the EDSS obtained by physical examination and a score derived from a self-assessment questionnaire using a special computer program. In an Italian multicentre study, self-assessment was compared with physical examination using the minimal record of disability (MRD) for MS and the EDSS.[10] Strong agreement was described especially for more disabled patients (ICC = 0.75). In the present study, overall agreement was better (ICC = 0.95) even for lower EDSS scores (ICC = 0.79). The correlation reported for the single functional system in patients with an EDSS score < 4.5 was similar to the present results, although they had especially poor values for Cerebellar FS (ICC = 0.03) and Sensory FS (ICC = 0.11).

In the Italian study,[10] patients with major cognitive deficits were excluded because cognitive impairment was assumed to interfere with the accuracy of self-assessment.

This problem was addressed here by comparing patients with cognitive deficit, as defined by a baseline FS score of > 1 in the Mental FS, with patients with no impairment, resulting in a value of zero in the Mental FS. No difference between the two groups regarding agreement on EDSS (ICC = 0.95 and 0.93 respectively) was found. However, it has to be taken into account that only one patient with severe dementia was included.

Taking the physical EDSS as gold standard, a high Pearson correlation was found between the physical

examination and the telephone interview regarding the endpoint EDSS (0.97), which indicates a linear relationship between the two tests. A plot of the mean values of telephone and physical assessment reveals a random scatter, which suggests that there is no bias in one way or the other (Figure 2). The interclass coefficient was 0.95. If only patients with an EDSS score of < 4.5 were used for calculation, then the measures for agreement are less: $R^2 = 0.64$, ICC = 0.79. This was expected because the number of cases was reduced. However, the plot reveals a random pattern. The cause of the lower values for EDSS score < 4.5 is that as the EDSS values are restricted, there is less systematic variation between cases to explain. Hence the error term dominates.

These results compare favourably with previous results of physical EDSS inter-rater reliability studies.[11] The values can not be expected to be higher than the inter-rater correlation between two physical examiners. The kappa value for full agreement in the current study was 0.48. Amato *et al.*[11] tested the inter-rater agreement on MS patients with an EDSS score ranging from 1.0 to 8.5 and found a total agreement of 0.5. Referring to a score difference of half a point (one-step variation), the kappa value increased to 0.75, whereas in the present study it was 0.79 and for a score difference of a full point (two-step variation) Amato *et al.* described a kappa of 0.96 compared with a value of 0.90 here.

The total agreement in the present questionnaire, especially for EDSS scores below 4.5, is only moderate (kappa = 0.24), which might be explained by the lack of detailed questions to assess single FS and the inability to assess EDSS scores of 1.0 and 1.5 by telephone. Verdier-Taillefer *et al.*[12] found in their study of 59 MS patients a kappa of 0.15–0.19 for an EDSS score < 5 and 0.61 for an EDSS score > 5.

Goodkin and colleagues[13] examined 10 clinically stable MS patients at three different times by four different neurologists. One hundred per cent agreement was only found in 10% for EDSS scores between 1.0 and 3.5. Francis *et al.*[14] compared three independent examiners rating 20 MS patients with EDSS scores from 3.0 to 9.0. The kappa values for EDSS were 0.32–0.76 depending on which pair of examiners was compared. Total agreement was 0.45; taking a two-step variation into account it increased only to 0.85. Noseworthy *et al.*[15] examined MS patients with a higher EDSS score (4.0–6.5) and found a better total agreement of 69%. The ICC of the present study population increases to 0.87 if only patients with an EDSS score > 4.5 are considered. This reflects again that the correlation is better when the values do not depend on too many variables (like FS) but mainly on walking distance.

Two recent studies[16,17] have also examined the psychometric properties of the EDSS in comparison with a couple of other scores. Sharrack *et al.*[16] assessed 64 patients with a wide range of disabilities by three different raters. They achieved substantially higher kappa values and ICC for the single FS (ranging from 0.41 to 0.67 and from 0.81 to 0.95, respectively). The inter-rater agreement on the EDSS scores was 96% when accepting a difference of < 1.0 points, which is similar to the present result (90%). Hobart *et al.*[17] examined 125 moderate to severely disabled MS patients by 11 different raters. Their ICC was 0.78 for the EDSS and a range of 0.38–0.72 for the single FS. These results are also similar to the present results with an ICC of 0.87 for EDSS scores > 4. Compared with the older studies, these two studies seem to achieve better agreement, which might be due to the fact that some of their raters participated at training workshops.[4,16,17]

Ambulation is a value expected to vary considerably between self- and observer-assessment. Sharrack and Hughes[18] compared the estimated walking distance from 100 patients with the estimation of their consultants. The estimates differed up to 14.6-fold from the measured distances. The performance depends heavily on patient's motivation. But there is also a variable degree of persistence by the examining neurologist. The patient's physical fitness varies not only from day to day but also between morning and afternoon. The acceptance of assistance varies also from patient to patient. Some MS patients choose to use a cane to avoid appearing drunk, due to an ataxic gait. Others will try to avoid the assistance of a cane until they are almost in need of a wheelchair. These contingencies result in inter-rater reliability problems for higher EDSS scores.

Timed walking tests usually show a higher reliability (correlation coefficient > 0.95) but still hide considerable individual patient variability of the order of 20%.[19]

In the questionnaire there is no zero value for the assessment of FS. Therefore, the agreement on FS which had values in the physical assessment of mainly 0, 1 and 2 was particular low; for example, visual impairment (kappa = 0.22). Surprisingly high values for Brainstem (0.59) and Pyramidal FS (0.54) were found, although in the clinical assessment these FS require most skill and training to achieve a high retest consistency.[4] There is some limitation of the study regarding the assessment of FS by telephone interview. The recoding of the physical values 0 and 1 might have inflated the ICC. At the same time, it decreases the agreement of the lower EDSS scores because of the inability to assess EDSS scores of 1.0 and 1.5 by telephone in most FS. Although both the Bowel/bladder and Mental FS rely also on questions in the physical examination, the agreement is lower than expected. The telephone questionnaire unfortunately does not take into account any bowel disturbances. In addition, the values of both FS vary less than the values for Pyramidal or Brainstem FS, which might explain partly the low ICC.

In one of the first reports to describe the inter-rater reliability in assessing FS and the EDSS,[11] the Visual FS was not included in the evaluation, avoiding the particular difficulties in this function. Lowest agreement was described for Pyramidal FS (0.28), Mental and Sensory (0.32), Brainstem and Bowel/bladder (0.5) and best values for Cerebellar FS (0.56).

In the series of Verdier-Taillefer *et al.*,[12] the inter-rater agreement for single FS was best for Cerebellar and Bowel/bladder with a kappa value of 0.53. Lowest values were found for Sensory 0.23 and Cerebral FS 0.25. In the study

of Francis *et al*.,[14] FS agreement was lowest for Cerebellar 0.14 and highest for Bowel/bladder 0.56. A better agreement was found for the Ambulation Index 0.5–0.7.

The structured telephone interview 'EDSS by telephone' is a valid tool to assess the EDSS in cases where the patient is unable to continue visiting the study centre. The telephone interview-derived scores are unbiased estimates for the cases studied for the EDSS scores 0–7, and may be biased by about 0.25 downwards for scores 8 and 9. This last result has to be confirmed with a larger number of patients, as only two patients with an EDSS score of 8 and two patients with an EDSS score of 9 were included. The reliability of the telephone EDSS is poorer for values less than 2, even though being unbiased. Better reliability was found in higher EDSS scores, where the assessment by telephone interview might be needed most. The kappa values for the EDSS and most FS were equivalent to or better than previous data of inter-rater reliability of the Kurtzke Scale.[11–17]

With the present study, sufficient inter-rater reliability of the 'EDSS by telephone' could be documented for physicians trained with a specific audiovisual training program.[4,20] The tool is not yet validated for use by untrained physicians or nonmedical professionals.

A telephone interview will not replace the physical examination by neurologists. But in clinical studies where regular visits are necessary and baseline data exist, assessment via telephone is a helpful and, as shown here, a reliable substitute if the patient is unable to attend a visit.

Acknowledgements

The authors thank Schering AG (Berlin) for the financial support of this study, V Rohrer for expert secretarial assistance, Prof. C Pozzilli (Rome) and K Wagner (Berlin) for valuable comments. Prof. L Kappos is supported by the Swiss MS Society.

References

1 Kurtzke JF. Rating neurological impairment in multiple sclerosis: an Expanded Disability Status Scale (EDSS). *Neurology* 1983; **33**: 1444–52.
2 Rudick R, Antel J, Confavreux C, Cutter G, Ellison G, Fisher J *et al*. Recommendations from the National Multiple Sclerosis Society Clinical outcome assessment task force. *Ann Neurol* 1997; **42**: 379–82.
3 Willoughby EW, Paty DW. Scales for rating impairment in multiple sclerosis: a critique. *Neurology* 1988; **38**: 1793–98.
4 Lechner-Scott J, Huber S, Kappos L. Expanded Disability Status Scale (EDSS) training for MS-multicenter trials. *J Neurol* 1997; **244**(suppl 3): 25.
5 Sanchez MM, Binkowitz BS. Guidelines for measurement validation in clinical trial design. *J Biopharm Stat* 1999; **9**: 417–38.
6 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–73.
7 Lublin F, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology* 1996; **46**: 907–11.
8 Weinshenker BG, Rice GPA, Noseworthy JH, Carriere W, Baskerville J, Ebers GC. The natural history of multiple sclerosis: a geographically based study. 1. Clinical course and disability. *Brain* 1989; **112**: 133–46.
9 Goodin DS. A questionnaire to assess neurological impairment in multiple sclerosis. *Mult Scler* 1998; **4**: 444–51.
10 Solari A, Amato MP, Bergamaschi R, Logroscino G, Citterio A, Bochicchio D, Filippini G. Accuracy of self-assessment of the minimal record of disability in patients with multiple sclerosis. *Acta Neurol Scand* 1993; **87**: 43–46.
11 Amato MP, Fratiglioni L, Groppi C, Siracusa G, Amaducci L. Interrater reliability in assessing functional systems and disability on the Kurtzke scale in multiple sclerosis. *Arch Neurol* 1988; **45**: 746–48.
12 Verdier-Taillefer MH, Zuber M, Lyon-Caen O, Clanet M, Gout O, Louis C, Alperovitch A. Observer disagreement in rating neurologic impairment in multiple sclerosis: facts and consequences. *Eur Neurol* 1991; **31**: 117–19.
13 Goodkin DE, Cookfair D, Wende K, Bourdette D, Pullicino P, Scherokman B *et al*. and the MS Collaborative Research Group. Inter-intrarater scoring agreement using grades 1.0 to 3.5 of the Kurtzke Expanded Disability Status Scale (EDSS). *Neurology* 1992; **42**: 859–63.
14 Francis DA, Bain P, Swan AV, Richard AC, Hughes AC. An assessment of disability rating scales used in multiple sclerosis. *Arch Neurol* 1991; **48**: 299–301.
15 Noseworthy JH, Vandervoort MK, Wong CJ, Ebers GC and the Canadian Coopertive MS Study Group. Interrater variability with the expanded disability status scale (EDSS) and functional systems (FS) in a multiple sclerosis trial. *Neurology* 1990; **40**: 971–75.
16 Sharrack B, Hughes RA, Soudain S, Dunn G. The psychometric properties of clinical rating scales in multiple sclerosis. *Brain* 1999; **122**: 141–59.
17 Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. *Brain* 2000; **123**: 1027–40.
18 Sharrack B, Hughes RA. Reliability of distance estimation by doctors and patients: cross sectional study. *BMJ* 1997; **315**: 1652–54.
19 Collen FM, Wade D. Assessing motor impairment after stroke: a pilot reliability study. *J Neurol Neurosurg Psychiatry* 1990; **53**: 567–79.
20 Kappos L, Lechner Scott J, Lienert C. Neurostatus training CD-ROM for a standardized neurological examination and assessment of Kurtzke's functional systems and expanded disability status scale for MS patients. Point de vue, Basel 1999, 2002, available through the authors.