

J.F. Kurtzke

Natural history and clinical outcome measures for multiple sclerosis studies. Why at the present time does EDSS scale remain a preferred outcome measure to evaluate disease evolution?

If one person can be considered the founder of modern clinical neurology, it would be Jean-Martin Charcot. His lectures at la Salpêtrière on diseases of the nervous system were given from 1866 to 1880. His illustrations of the spinal cord in multiple sclerosis (MS) show the high proportion of white matter involved by the lesions which are "scattered in time and space." Thus, the major pathology in MS is concentrated in the neural tube. The principal clinical deficits in MS then would be the result of lesions in the spinal cord and brain stem, with multiple lesions along the neural pathways required for impairment of function. Therefore, the predominant involvement of long tracts would most often be reflected in the lower limbs, and bilaterally.

Presented at the International Workshop. Highlights on Multiple Sclerosis, 7 April 2000, University of Ferrara, Ferrara, Italy

J.F. Kurtzke (✉)
Neuroepidemiology Section (127)
Veterans Affairs Medical Center
50 Irving Street, NW
Washington, DC, 20422 USA

The early clinical features in MS were studied in a series of 762 men with final diagnosis of MS in US Army Hospitals during World War II. Neurologic exams during and after military service for the first 20 years of illness were obtained. For each exam, symptoms were recorded by type and body part affected, and signs were coded to the eight Functional Systems (FS) and the Disability Status Scale [1]. Two-thirds of the patients were considered to have definite or probable MS.

All these patients were free of neurologic impairment before the bout leading to Army hospital diagnosis. In that bout, 85% then had Pyramidal tract involvement. Cerebellar, Brain Stem and Sensory systems were the other Functional Systems involved in the majority of patients even at this very early stage of clinical disease. Together with the four other systems (Bowel and Bladder, Visual, Cerebral, Other), these 8 Functional Systems – which are all mutually exclusive – include all deficits seen in this disease that can be defined at neurologic examination.

With this exclusivity and completeness, each patient can then be described by an 8-digit binary number, with a "1" for involvement in a given system, and a "0" for no involvement. For example, 1100 0000 would describe a patient with only Pyramidal and Cerebellar signs. Now if occurrence of disease affecting one FS is independent of involvement in any other FS, then a specific frequency expected for each of the individual 256 possible patterns (2^8) can be calculated as the product of each of the observed frequencies within the total series. For involvement of only Pyramidal and Cerebellar Systems, this expected frequency would be the product of observed values (the "1s") and that of all six "0s" ($1 - \text{observed values}$). This product is .028, and the observed product was .024. This pattern was the 6th most common expected. The four most common were 1111 0000, 1110 0000, 1111 0100, and 1110 0100, which together comprised 26% of cases expected – and observed. Fourteen specific patterns were expected to constitute half of all the cases, and they did. Ninety percent of patients were expected to fall within the 86 most common patterns, and 90% were observed. Most patients showed patterns involving chiefly

the four major Functional Systems, and multiple minor system involvement was rare.

The Functional Systems, however, are of limited value in describing patterns of neurologic involvement when each system is quantified. There would then be several million possible patterns into which patients might fall. Adding scores for the separate systems is really not valid, as these are not arithmetic scales, and steps in one scale are not equivalent to those in another. When this was done, though, values plateaued well below possible maxima. There are also patients who over time improve in one area while they worsen in another – plus the interplay of Pyramidal and Cerebellar functions. These are the reasons why an overall measure of neurologic involvement was devised in order to measure change in MS, and this was the Disability Status Scale (DSS). The DSS later evolved into the Expanded DSS by dividing into two each step from DSS 1 through 9.

In the Army hospital series, some 1700 examinations in the first 20 years of illness among the MS patients were scored on the DSS. Overall, they showed a unimodal and fairly normal distribution. The frequency of involvement in each Functional System increased with increasing DSS scores, but the order of involvement, Pyramidal more than Cerebellar more than Brain Stem, and so on, remained almost the same regardless of DSS. For the more severely involved patients, almost all had Pyramidal and Cerebellar signs. Only Bowel and Bladder Function altered the pattern, with even more frequent involvement for this spinal cord sign than for some other FS in the more severe patients.

Comparing each FS by step versus the DSS in this series showed for each of them an increasing frequency *and* severity of involvement. In the order listed, the frequency distributions for all functional systems had a progressively increasing shift to the left. This would indicate lesser degrees of involvement qualitatively and quantitatively as we go from pyramidal to cerebral. Basing an overall scale strongly on ambulation would thus appear to be an asset rather than a defect, since it properly reflects how most patients are involved clinically in this disease – and at all stages of the illness.

Relationships among the functional systems and the DSS were explored with Tschuprow coefficients of correlation. This is a non-parametric measure between two variables, with perfect agreement as 1.00 and absence of agreement as 0. Each FS was significantly correlated with every other one, except for Cerebral versus Bowel and Bladder, again indicating the interrelationships of these measures of neurologic impairment. Strongest correlations were those with the DSS in each instance, with the possible exception of the high Bowel and Bladder versus Pyramidal coefficient.

DSS scores in the Army series were also used to investigate prognosis. Bout frequency early in the disease has been used as one such measure. While there was a slight trend towards worse disease with increasing bouts, it did not seem a very strong predictor. In like manner, the severity of neurologic impairment at first diagnosis did not appear too use-

ful. However, the neurologic status at five years after onset was a very strong predictor of the course for the next 13 years. Two-thirds of patients with DSS 0-2 at five years after onset were still 0-2 at 15 years after diagnosis. For severely involved patients (DSS 6-9) at five years, almost all were severe at the latter point, and half were dead. This has been called by others the “five-year rule.”

While there are then a number of uses for the DSS and FS, the origin of these scales was as a means of assessing results in a treatment trial. Our early findings with isoniazid appeared striking when compared with prior experience at the same hospital as measured by change in DSS scores between admission and discharge [2]. This work led to the first multicentered, randomized, placebo-controlled, double-blind, therapeutic trial ever carried out in MS. The trial was totally negative as to treatment efficacy, and a 5-year follow-up confirmed the lack of effect. Our own later experience had also led to the same conclusion, but also to the conclusion that one *could* evaluate therapy in this disease, but only with a proper double-blind methodology with placebo control.

The next step in therapeutic trials in MS in the USA took place almost a decade later. This was a randomized trial of short-term ACTH versus placebo in acute bouts of MS in ten University-based centers [3].

Assessment methods were several. One was a standardized neurologic examination with 39 items, each graded on a 5-point scale from 0 (none) to 4 (total), plus two items for visual acuity (0-8). Points were added and divided by 39 (or 41) for the exam score. Next were the Functional Systems. Scoring here was as the sum of the grades in each of the seven scaled systems, divided by seven. Devised for this study was the Seven Day Symptom Score. The presence and severity (3 steps) of 20 specified neurologic symptoms were recorded for each of the seven days preceding the exam. Severity times duration divided by 20 gave the score for that week. Tourtellotte’s quantitative method (QENF) was also used. This is a combination of standardized timed motor and coordination tests plus quantitated sensory exams. Average exam differences versus baseline were each normalized and summed, and then divided by 50 (for the number of items tested) in order to provide the score for that exam. Since these reflect deviations from normal, all scores are negative in reference to the other methods – and to normal function.

The other methods used were the EOC and the DSS. The pretreatment overall impression of neurologic status (mild, moderate, severe) became the estimate of overall condition (EOC). Here the physician and patient agreed as to whether he was better, worse, or the same at each weekly exam versus the pretreatment baseline condition, without documenting the nature of any change. This is valid so long as a study is truly double-blind as to treatment. Concordance of all measures of pretreatment severity was good, with very high coefficients of correlation, even with this parametric testing. DSS was possibly a bit better than the others.

On the EOC there was a significant difference favoring treatment, which seemed to lessen with time. Over 80% of patients improved on ACTH with this measure, while the placebo group rose with time to nearly 70% improved. On the DSS, though, the difference favoring treatment persisted throughout the study. At the end 65% of treated patients improved on the DSS versus 48% of the placebo group - remarkably close to predictions. Results were quite uniform among the centers. Whether ACTH made a difference in the course or merely hastened improvement then was a decision based on which method of measurement was chosen.

A similar comparison for each Functional System was carried out. Changes in each FS were strongly correlated with DSS changes, and notably less with the EOC. Correlations among FS were significant within most of the four major systems but largely not so among the minor systems. This probably reflects the low frequencies of involvement for the latter. For the study as a whole, all the methods of evaluation used correlated about equally well with one another in terms of measures of improvement. The DSS, FS and standard neurologic exam were perhaps the strongest.

The system of EDSS plus FS has also been used in a number of more recent treatment trials. The EDSS has also found a role in characterizing MS in some other descriptive and natural history studies. As to the evaluation of treatment,

there are certainly other methods than those considered here. For such trials, though, I believe we need to remain with measures of neurologic impairment. It is likely that any method that gives worse scores for worse neurologic condition, and the reverse, would likely be usable in treatment trials, so long as they are properly designed and carried out. The reason is that, as we have seen, as groups of patients change neurologically, they tend to change in all testable areas. However, as of now, I believe that the method most accurately reflecting such states and changes, is still that of the EDSS and Functional Systems.

References

1. Kurtzke JF (1989) The Disability Status Scale for multiple sclerosis: Apologia pro DSS sua. *Neurology* 39:291-302 [other works cited are referenced here]
2. Kurtzke JF, Berlin L (1954) The effects of isoniazid on patients with multiple sclerosis. Preliminary report. *Am Rev Tuberc* 70:577-592
3. Rose AS, Kuzma JW, Kurtzke JF, Namerow NS, Sibley WA, Tourtellotte WW et al (1970) Cooperative study in the evaluation of therapy in multiple sclerosis: ACTH vs. placebo. Final report. *Neurology* 20 (No 5, Part 2):1-59