Movement Disorders
Vol. 12, No. 2, 1997, pp. 183–189

The Need for Levodopa as an End Point of Parkinson's Disease Progression in a Clinical Trial of Selegiline and α-Tocopherol

Peter LeWitt, *David Oakes, *Lu Cui, and the Parkinson Study Group

*Clinical Neuroscience Center, Sinai Hospital, West Bloomfield, and the Department of Neurology and the Clinical and Cellular Neuroscience Program, Wayne State University School of Medicine, Detroit, Michigan; and *Department of Biostatistics, University of Rochester School of Medicine, Rochester, New York, U.S.A.*

Summary: Progression of Parkinson's disease (PD) can be detected through changes in clinical ratings or disability assessments. A clinical trial, Deprenyl and Alpha-Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP), used a novel study end point: increase in parkinsonian disability to the extent that investigators determined the need for treatment with levodopa. We analyzed DATATOP results to learn if this operationally defined end point could be reproduced from elements of the Unified PD Rating Scale (UPDRS) and other conventional clinical scales. Our analysis involved UPDRS, Schwab and England Activities of Daily Living (S-E ADL), and Hoehn andYahr (H-Y) scores when DATATOP subjects reached the study end point. Various UPDRS components were examined, including subscores measuring severity of impaired ADL, bradykinesia, postural instability and gait difficulty, tremor, and rigidity. Data from subjects reaching the end point were compared with assessments of those DATATOP subjects who did not, matched for the same duration of enrollment. All measures showed subjects who reached the end point had significantly greater mean impairment than did controls, although the two groups had substantial overlap. Multivariate analysis by using conditional logistic regression suggested that the end point was determined more by functional (S-E ADL and the UPDRS ADL scores) than by clinical examination criteria. The method of classification and regression trees suggested a simple decision tree splitting, respectively, on S-E ADL, UPDRS ADL, H-Y score, and UPDRS ADL again, with an estimated overall misclassification probability of 18%. We conclude that the DATATOP end point cannot be fully reproduced from the traditional clinical measures, although it can give results that are consistent with these scales in a well-designed clinical trial.
Key Words: Clinical trial—DATATOP—Parkinson's disease—Primary end point—Selegiline—Rating scales.

A variety of rating scales have been developed for evaluating the severity of parkinsonian signs, symptoms, and disabilities (1–4). These measures have been used primarily for gauging improvement from medical and surgical treatments, although they also can be used for describing severity and assessing change in the underlying parkinsonian disorder. For ~10 years, the Unified Parkinson's Disease Rating Scale (UPDRS) (4,5) has served as a descriptive and semiquantitative instrument for use in clinical studies. This scale comprises 31 questions describing mentation, mood, behavior, activities of daily living (ADLs), and parkinsonian motor impairments, each rated for severity by using a 5-element scale. Although additional assessments can be useful for evaluating other aspects of parkinsonism [such as the pattern of responsiveness to levodopa (6)], the UPDRS gives a thorough description of Parkinson's disease (PD) and has been studied for its validity and internal consistency (7–9).

An interest in interventions against the disease mechanism in PD led to the design of a clinical trial in which the evolution of parkinsonian disability was monitored (10). This trial, entitled Deprenyl and Alpha-Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP) (10,11), used a novel determinant of PD progression as the primary study end point. This operationally defined end point was reached when the clinician-investigator decided that a subject's initially mild parkinsonism had progressed to the extent that symptomatic treatment

Received May 4, 1995, and in revised form January 29, 1996. Accepted May 13, 1996.

Parkinson Study Group members are listed in references 10, 11, and 16.

Address correspondence and reprint requests to Dr. P. A. LeWitt at Clinical Neuroscience Center, 5821 West Maple Road, Suite 192, West Bloomfield, MI 48322, U.S.A.

(levodopa or its equivalent) became necessary for treating disability. Although first used in the DATATOP study and in a smaller parallel study (12), this end point subsequently was adopted for other clinical trials (13,14). In the DATATOP study, emergence of difficulties in the workplace, disabilities in ADLs, or worsening of gait and balance were major factors that influenced the need for starting levodopa therapy (11). This trial also collected data from the UPDRS and other tests quantifying parkinsonian features, thereby permitting a thorough characterization of subjects before and upon reaching the study's end point.

Knowledge of the interval between randomization and the determination of end point (the "survival" interval) permitted statistical analysis by the method of Kaplan and Meier for the cumulative probabilities of reaching the endpoint (15) after treatment with placebo, selegiline (deprenyl), α-tocopherol, or their combination. As described in an interim report in 1989 (11), this trial indicated that selegiline attenuated the rate at which subjects reached end point during the first 12 months of treatment. Selegiline treatment also reduced, by approximately half, the rate of worsening in various UPDRS subscores. These prominent effects in the first year continued but did not increase during the subsequent 6 months of follow-up (16). α-Tocopherol treatment had no effect.

Whereas the results from the DATATOP trial have been subjected to rigorous analysis (23), the subjective and complex nature of end point determinations engendered concerns that factors other than progression of PD might have exerted some influence. For example, a subject's desire to remain in the trial despite emerging disability might have delayed the determination of an end point, just as a subject's unwillingness to continue in the trial and endure a stable parkinsonian state might have precipitated a premature end-point determination. The randomized, double-blinded study design made such possibilities unlikely to confound the comparisons made between the four treatment groups (placebo, selegiline, α-tocopherol, and selegiline plus α-tocopherol). Each investgator had aproximately equal numbers of subjects receiving the four treatments, and data analyses were stratified by investigator (so that any personal bias in the determination of end point might be applied in a similar manner to subjects on each of the treatments). Analyses of the UPDRS changes showed the same striking effect of selegiline against progression of parkinsonism as was revealed by the primary end point analysis. Nevertheless, it is of interest to learn to what extent the decision for declaring a subject's end point at a particular visit might have been predicted from the subject's clinical status at that visit as indicated by the UPDRS, Schwab and England ADL, and Hoehn and Yahr staging. Another question we raised was whether selegiline treatment exerted an influence on the decision to declare end point independently of its effects on the various measures of parkinsonism and its disabilities.

MATERIALS AND METHODS

Subjects

In the DATATOP study, 800 subjects between the ages 30 and 79 years and with mild idiopathic PD lasting ≤5 years were enrolled at 28 centers by 34 investigators (10). At entrance into the study, subjects were determined to be free of significant overall parkinsonian disability and were willing to forego levodopa or other symptomatic therapy until the advent of significant disability. Factors for exclusion included the presence of dementia, significant depression, or resting tremor of ≥3 on the UPDRS scale, which ranged from 0 (none) to 4 (maximal severity) (5). If subjects had previous treatment with levodopa, selegiline, or other symptomatic therapy, these medications had to be discontinued before participation. At entry into the DATATOP study, the mean duration of parkinsonian symptoms (when known, as was the case for 748 subjects), was 2.1 ± 1.3 years, and the mean age was 61.1 ± 9.5 years. Further demographic data pertaining to participants have been published (10,11,16).

Subjects were randomly assigned to one of four treatment groups for continued administration of either selegiline (10 mg/day), α-tocopherol (2,000 IU/day), or their placebos: (a) α-tocopherol–placebo plus selegiline–placebo; (b) α-tocopherol plus selegiline–placebo; (c) selegiline plus α-tocopherol–placebo; and (d) α-tocopherol plus selegiline. The randomization process was stratified to achieve an approximate balance of the number of subjects enrolled by each investigator into the four treatment groups.

Ratings

At all study visits, the same investigator performed UPDRS ratings for each enrolled subject (5). Regular clinic visits were scheduled at 1 and 3 months after randomization, and at ~3-month intervals thereafter. A subject's ratings could be carried out at other times if disability evolved before the next scheduled visit.

The primary end point of the study occurred when the enrolling investigator determined that a subject had reached a level of functional disability sufficient to warrant the initiation of levodopa therapy. At each clinic visit, subjects were reviewed as to whether they had reached end point criteria, and assessments were con-

ducted of UPDRS, Schwab and England ADL, and Hoehn and Yahr staging. The version of the UPDRS used in the DATATOP study was modified to permit half-point intervals between rating categories on the motor examination.

Analytic Methods

For the purpose of analyses to be reported here, UPDRS ratings (4,5) were grouped as follows:

Mental Symptom assessments (the sum of questions 1–4);

Activities of Daily Living assessments (the sum of questions 5–17);

Clinical Motor Examination (the sum of questions 18–31); and

Total UPDRS score (the total of Mental Symptoms, ADL, and the Clinical Motor Examination assessments).

Scores on the Schwab and England Activities of Daily Living (S-E ADL) (17) (modified to permit determinations at 5 percentile–unit increments) and Hoehn and Yahr (H-Y) scales (18) (modified to permit determinations of half-point increments) were also studied. Another evaluation in this study was a timed task of manual dexterity, the Purdue pegboard test, which involved counting the number of small pegs that a subject could insert bimanually into a row of holes drilled into a board (19).

Further analysis was carried out with various summed groupings of motor ratings, categorized in a previous study (20) as tremor, rigidity, bradykinesia, and the Postural Instability and Gait Difficulty (PIGD) score:

Tremor Score (the sum of all Clinical Motor Examination tremor rating items);

Rigidity Score (the sum of all Clinical Motor Examination rigidity rating items);

Bradykinesia Score (the sum of all Clinical Motor Examination bradykinesia items);

Postural Instability and Gait Difficulty (PIGD) Score (the sum of UPDRS items pertaining to falling, freezing, history of walking difficulty, and gait and postural instability on clinical examination).

The previous study (20) used the ratio of the PIGD score to the tremor score to differentiate PIGD patients from tremor-predominant patients—in this analysis, we included the numerical values of the calculated subscales.

In our study, the plan was to contrast data derived from the UPDRS and other assessments from subjects who reached end point (end-pointed subjects) at a particular visit with data from other subjects who remained active beyond that visit. End-pointed subjects were matched with three "active" subjects who had the same gender and were enrolled by the same investigator. For the purposes of this analysis, the active controls chosen had to "survive" without reaching end point until at least the next regularly scheduled visit.

Distributions of each variable among cases and controls were tested by simple t tests to determine how well each discriminated between the two groups. Then a multivariate analysis was conducted with stepwise matched logistic regression (21) to determine which combination of variables was best at discriminating subjects who reached end point from those remaining as active controls.

Two stepwise conditional regression analyses (21) were performed by using the EGRET program (Statistics and Epidemiology Research Corporation, Seattle, WA, U.S.A.). The first analysis involved subject age, the three major groupings of the UPDRS (the summed Mental Symptom, ADL, and Clinical Motor Examination items), the S-E ADL score, the H-Y stage, and the treatment assignment of whether or not selegiline was administered. In the second analysis, each of these variables was studied with the summed tremor, rigidity, and bradykinesia components of the UPDRS. Numeric variables were entered without dichotomization.

The second method of analysis (22) used the method of Classification and Regression Trees (CART). This procedure constructed a sequence of binary questions at each node, examining all possible cut-points of all possible variables to derive a single dichotomy that best separated the visits at that node into end-pointed and active status. Cross-validation was used to assess the utility of including additional nodes (i.e., the size of the tree was determined by the ability of classification rules derived from one randomly selected subset of the data to classify correctly the observations not included in the subset used to derive the classification rules). The CART analysis included the same variables as the first conditional logistic regression analysis.

RESULTS

Frequency Distributions

Table 1 shows means and standard deviations of the various assessments for subjects reaching end point and for their matched active controls. In all cases, subjects reaching end point were more impaired on average than were active subjects. However, there was considerable overlap between the distributions—this is illustrated graphically by histograms of UPDRS (total motor score) and Schwab and England ADLs (Fig. 1). Table 1 also

TABLE 1. *Means and standard deviations (or percentages, as appropriate) of selected clinical measures at 375 end point vitis and a matched comparison sample of 1,041 "active" visits*

	End point visit (n = 375)	"Active" visit (n = 1041)
End point values		
Age	60.70 (9.30)	61.50 (9.30)
Hoehn and Yahr staging	2.24 (0.52)	1.70 (0.51)
Schwab and England ADL	78.50 (9.70)	89.90 (7.10)
UPDRS		
Total	45.60 (14.10)	27.70 (13.40)
Motor	29.90 (10.50)	18.80 (9.90)
Mental	2.00 (1.80)	1.00 (1.20)
ADL	13.70 (4.50)	7.90 (4.10)
Tremor	7.20 (4.40)	5.00 (3.40)
Rigidity	6.30 (3.40)	4.10 (3.10)
Bradykinesia	1.80 (0.90)	1.00 (0.80)
PIGD	3.20 (2.10)	1.60 (1.30)
Purdue Pegboard	6.30 (2.90)	7.50 (2.80)
Selegiline treatment	153 (41%)	591 (57%)
Changes from immediately preceding visit[a]		
Hoehn and Yahr staging	0.29 (0.50)	0.06 (0.40)
Schwab and England ADL	6.20 (7.50)	0.70 (4.60)
UPDRS		
Total	8.90 (8.70)	2.10 (5.80)
Motor	5.60 (6.20)	1.50 (4.50)
Mental	0.50 (1.50)	0.10 (1.10)
ADL	2.80 (3.30)	0.50 (2.30)
Tremor	1.10 (2.50)	0.30 (1.80)
Rigidity	1.10 (2.10)	0.20 (1.70)
Bradykinesia	0.40 (0.60)	0.10 (0.50)
PIGD	1.00 (1.50)	0.20 (0.90)

ADL, Activities of Daily Living; UPDRS, Unified Parkinson's Disease Rating Scale; PIGD, Postural Instability and Gait Difficulty.

[a] All changes represent a worsening of clinical status in both groups.

shows changes in clinical status from the immediately preceding visit: again, relative worsening was much greater for end-pointed than for active subjects.

Stepwise Conditional Regression Analysis

The first stepwise conditional analysis yielded a selection of several variables, as shown in Table 2. In this analysis, a 1-point increase of S-E ADL, H-Y score, or UPDRS ADL resulted in a corresponding change (by the factor listed) for the odds of reaching end point. For example, an increase of 1 year in age decreased the odds for reaching end point by a factor of 0.96, whereas treatment with selegiline yielded decreased odds of reaching end point by a factor of 0.63.

The second stepwise conditional analysis produced similar results but also selected the tremor and rigidity components of the UPDRS as predictive of reaching end point (Table 2).

Classification and Regression Trees (CART)

The classification rules developed by the classification and regression trees (CART) program correctly classified 234 (62%) of 375 visits at which subjects reached end point, and 955 (92%) of 1,041 visits when other subjects remained active (Fig. 2). The overall misclassification rate with the use of CART, therefore, was (141 + 86)/(375 + 1,041), or 16%. However, because this finding was calculated with the same data used to develop the classification rule itself, this misclassification probability was probably too optimistic. The CART program also calculated a "cross-validated" rate by dividing the data into 10 groups at random and used the rule developed for each 90% of the group to classify the remaining 10%. The 10 misclassification rates were then averaged. By using this method, the "true" misclassification rate was then estimated to be 18%.

DISCUSSION

The need for starting symptomatic treatment is a distinctive, clinically important milestone in the progression of PD. For this and other reasons, the need for levodopa therapy, as determined by the same clinician-investigator who enrolled and monitored the patient, was selected as the primary end point in the DATATOP study. Although this end point has no simple equivalent in the UPDRS, its determination would be expected to be influenced by the

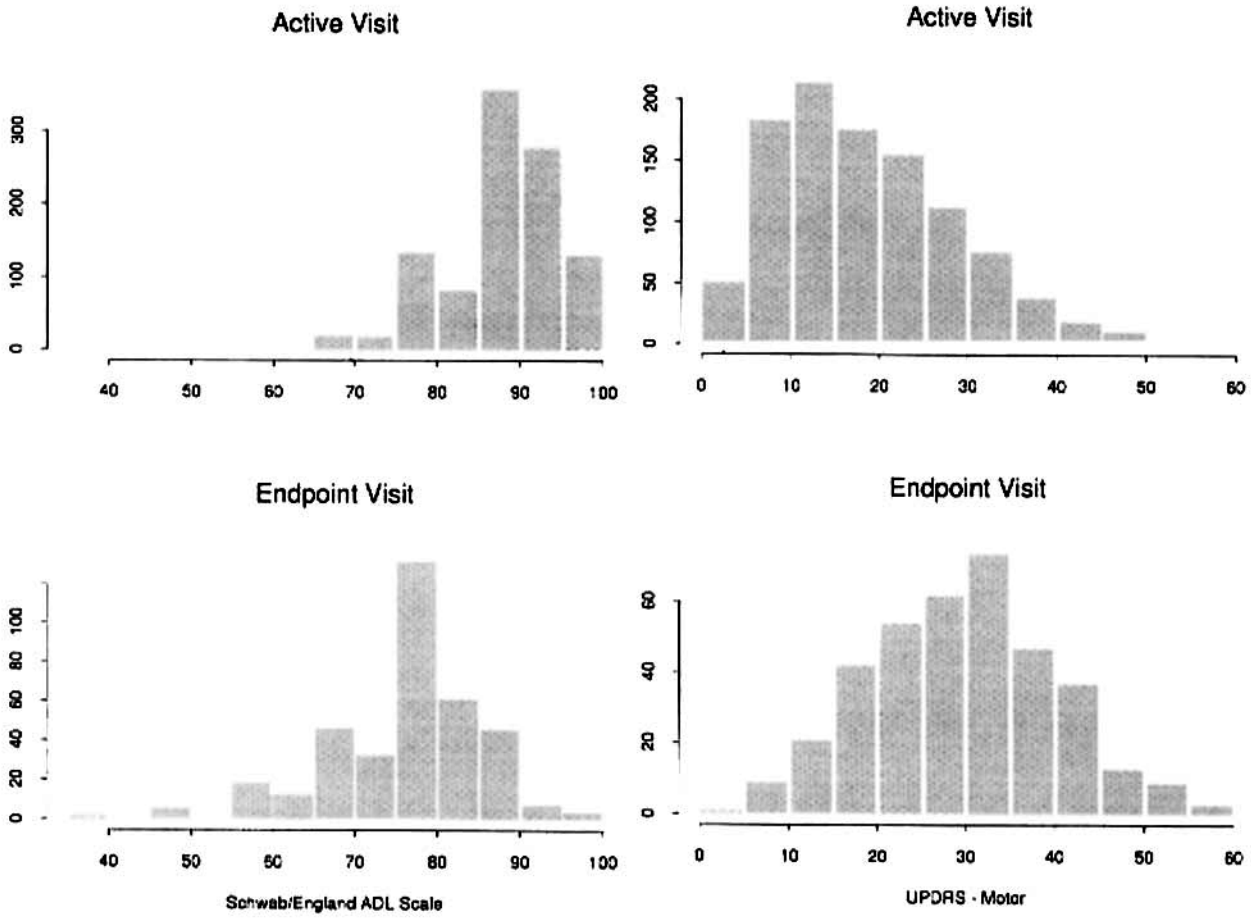

FIG. 1. Distribution histograms of numbers of subjects reaching end point (bottom) or still active in the study (top) for Unified Parkinson's Disease Rating Scale (UPDRS) and Schwab and England Activities of Daily Living (ADL) Scale rating (left) and sums of UPDRS motor rating components.

same factors involving disease progression as would lead to worsening as measured by the UPDRS and other clinical measures. However, no formal study has been carried out to assess the reliability or validity of "need for levodopa therapy" as a measure of disease progression, These considerations are important in view of the controversy that has arisen concerning the magnitude or mechanism of selegiline's effects.

The results of DATATOP showed a strong effect of selegiline both in delaying the primary end point and in reducing the rate of progression of illness, as measured by the UPDRS and its components, the Schwab and England

TABLE 2. *Results of the stepwise conditional regression analysis*

Variable	Coefficient	SEM	z-Value	Odds ratio[a]	p
First Analysis					
S-E ADL	−0.124	0.018	6.9	0.88	<0.0001
H-Y Score	1.30	0.240	5.4	3.65	<0.0001
UPDRS-ADL	0.136	0.030	4.5	1.15	<0.0001
Age	−0.0407	0.011	3.7	0.96	0.0002
Selegiline	−0.46	0.190	2.4	0.63	0.016
Second analysis					
Tremor	0.15	0.032	4.8	1.16	0.0001
Rigidity	0.18	0.055	3.2	1.20	0.001

S-E ADL, Schwab and England Activities of Daily Living score; H-Y score, Hoehn and Yahr score; UPDRS-ADL, Unified Parkinson Disease Rating Scale ADL score.

[a] Increase in odds per unit change in the score.

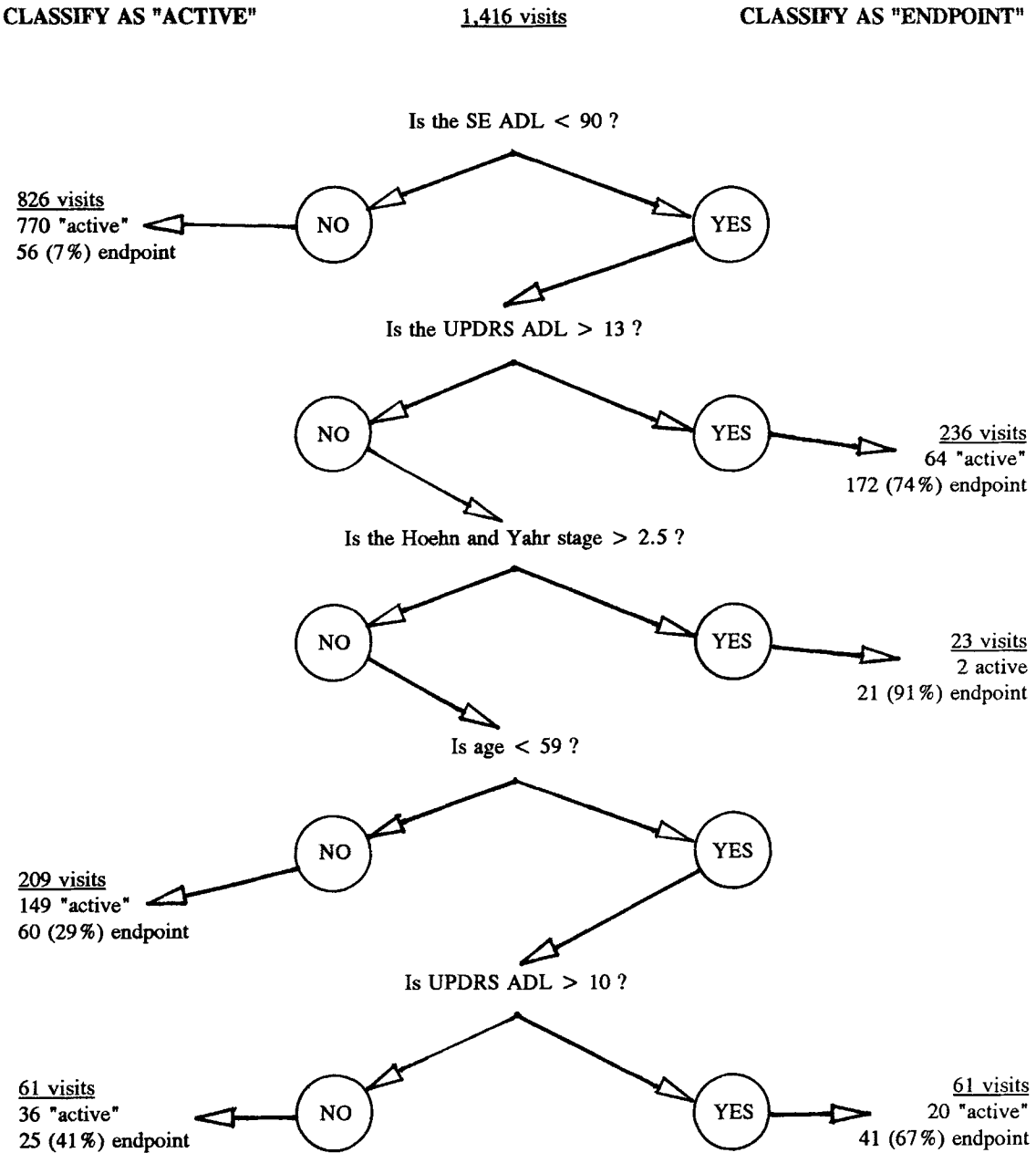

FIG. 2. Classification analysis of 1,416 Deprenyl and Alpha-Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP) visits (375 end point visits and 1,041 "active" visits) by using classification rules developed by the classification and regression trees program (see text). Variables included in the analysis: UPDRS (total, motor, mental, and ADL components), Hoehn and Yahr stage, Schwab and England ADL score, age, and assignment to selegiline treatment. Gender was not inclded because it was a matching variable.

ADL, and the Hoehn and Yahr staging. Moreover, patients who reached end point had progressed more rapidly (on average) on all measures than patients who did not reach end point. Analyses that used the primary end point cohered strongly with those that used the more conventional scales.

Our analyses confirm this coherence. Patients who reached end point were, on average, much more impaired than other patients who had completed the evaluations by the same investigator at the same interval after entering the study. On average, the end-pointed subjects had also shown a much greater decline since their previous visit

than their counterparts who did not reach end point. Our results also show that the measures, as applied to individual patients, are not interchangeable, although they do give similar results when applied to populations in well-designed clinical trials. Investigators and their patients differed in their assessment of the threshold of illness needed to prompt immediate symptomatic treatment. Substantial overlap between "end-pointing" and "surviving" subjects was found even for the more quantitative elements of the DATATOP evaluation battery, such as the Purdue Pegboard or other timed-movement scores. Some subjects not classified as reaching end point nonetheless had significant levels of disability (as indicated by their ADL scores).

Both the logistic regression analyses and the CART analyses showed that determination of end point was influenced more by measures of ADL than by motor ratings of parkinsonian features. One important difference between the two analyses was that the logistic regression analysis selected selegiline treatment as a predictor of survival even after adjustment for ADL and for clinical status. Selegiline delays end point in part by reducing rates of progression of illness and of decline in ADL; however, this treatment appears to have a further effect at delaying end point that was not captured by the various clinical measures used in these analyses. In a study as large as DATATOP, statistical significance is a weaker criterion than predictive ability—unlike the logistic regression, the classification analyses did not select selegiline as an independent predictor. Our finding concerning this possible independent effect of selegiline (if confirmed) indicate that rating scales cannot replace the DATATOP end point (or vice versa) and that both types of measure are useful and should be evaluated in clinical trials.

Acknowledgment: The DATATOP trial was supported by USPHS grant NS 24778. P.L. is supported in part by the National Parkinson Foundation.

REFERENCES

1. Teräväinen H, Calne D. Quantitative assessment of parkinsonian deficits. In: Rinne UK, Klinger M, Stamm G, eds. *Parkinson's disease—current progress, problems and management*. Amsterdam: Elsevier North-Holland, 1980:145–164.
2. Larsen TA, LeWitt PA, Calne DB. Theoretical and practical issues in assessment of deficits and therapy in parkinsonism. In: Calne DB, Horowski R, McDonald RJ, Wuttke W, eds. *Lisuride and other dopamine agonists*. New York: Raven Press, 1982:363–373.
3. Ward CD, Sanes JN, Dambrosia JM, Calne DB. Methods for evaluating treatment in Parkinson's disease. *Adv Neurol* 1983;37: 1–7.
4. Lang AE, Fahn S. Assessment of Parkinson's disease. In: Munsat TL, ed. *Quantification of neurological deficit*. Boston: Butterworths, 1989:285–301.
5. Fahn S, Elton RL, Members of the UPDRS Development Committee. Unified Parkinson's Disease Rating Scale. In: Calne DB, Fahn S, Marsden CD, Goldstein M, eds. *Recent developments in Parkinson's disease*. Vol 2. Florham Park, NJ: Macmillan, 1987: 153–163.
6. CAPIT Committee, Langson JW, Widner H, et al. Core assessment program for intracerebral transplantations (CAPIT). *Mov Disord* 1992;7:2–13.
7. Baas H, Stecker K, Fischer PA. Value and appropriate use of rating scales and apparative measurements in quantification of disability in Parkinson's disease. *J Neural Transm [P-D Sect]* 1993;5:45–61.
8. Martínez-Martín, Gil-Nagel A, Morlan Gracia L, et al., and The Cooperative Multicentric Group. Unified Parkinson's Disease Rating Scale characteristics and structure. *Mov Disord* 1994;9:76–83.
9. van Hilten JJ, van der Zwan AD, Zwinderman AH, Roos RAC. Rating impairment and disability in Parkinson's disease: evaluation of the Unified Parkinson's Disease Rating Scale. *Mov Disord* 1994;9:84–88.
10. Parkinson Study Group. DATATOP: a multicenter controlled clinical trial in early Parkinson's disease. *Arch Neurol* 1989;46: 1052–1060.
11. Parkinson Study Group. Effect of deprenyl on the progression of disability in early Parkinson's disease. *N Engl J Med* 1989;321: 1364–1371.
12. Tetrud JW, Langston JW. The effect of deprenyl (selegiline) on the natural history of Parkinson's disease. *Science* 1989;145:519–522.
13. Allain H, Cougnard J, Neukirch H-C, French Selegiline Multicenter Trial Members. Selegiline in de novo parkinsonian patients: The French Selegiline Multicenter trial. *Acta Neurol Scand* 1991;84(suppl 136):73–78.
14. Myllylä VV, Sotaniemi KA, Vourinen JA, Heinonen EH. Selegiline as a primary treatment of Parkinson's disease. *Acta Neurol Scand* 1991;84(suppl 136):70–72.
15. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:471–481.
16. Parkinson Study Group. Effects of tocopherol and deprenyl on the progression of disability in early Parkinson's Disease. *N Engl J Med* 1993;328:176–183.
17. Schwab RS, England AC. Projection technique for evaluating surgery in Parkinson's disease. In: Gillingham FJ, Donaldon IML, eds. *Third symposium on Parkinson's disease*. Edinburgh: Livingstone, 1969:152–157.
18. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology* 1967;17:427–442.
19. Cassell K, Shaw K, Stern G. A computerized tracking technique for the assessment of motor disabilities. *Brain* 1973;96:815–826.
20. Jankovic J, McDermott M, Carter J, et al. Variable expression of Parkinson's disease: a baseline analysis of the DATATOP cohort. *Neurology* 1990;40:1529–1534.
21. Breslow NE, Day NE. *Statistical methods in cancer research, Vol 1: the analysis of case control studies*. Lyon: IARC, 1980.
22. Breiman L, Friedman JH, Ohshen RA, Stone CJ. *Classification and regression trees*. Belmont: Wadsworth, 1984.
23. Parkinson Study Group. Impact of deprenyl and tocopherol treatment on Parkinson's disease in DATATOP subjects not requiring levodopa. *Ann Neurol* 1996;39:29–36.