Br. J. clin. Pharmac. (1991), **31**, 295–304 ADONIS 030652519100058K

Objective outcome criteria in trials of anti-Parkinsonian therapy in the elderly: sensitivity, specificity and reliability of measures of brady- and hypo-kinesia

S. G. BOWES[1], P. K. CLARK[2], A. CHARLETT[2], C. J. A. O'NEILL[1], A. L. LEEMAN[1,5], C. WELLER[3], P. W. NICHOLSON[1,4], A. A. DESHMUKH[1,5], S. M. DOBBS[1] & R. J. DOBBS[1]

[1]Therapeutics in the Elderly, Research Group, the [2]Section of Medical Statistics and [3]Division of Bioengineering, Clinical Research Centre, Northwick Park Hospital, Watford Road, Harrow, Middlesex HA1 3UJ, [4]University College and Middlesex School of Medicine, London W1P 6DB and [5]The School of Pharmacy, Brunswick Square, London WC1 1AX

1 We compare the sensitivity and specificity of chosen outcome criteria in a placebo-controlled, randomised cross-over study of the efficacy of maintenance therapy with the levodopa/carbidopa combination (Sinemet Plus) alone. Patients were characterised by having idiopathic Parkinsonism with no overt fluctuations in control in relation to individual doses of medication.

2 The effect of omission of a morning dose of maintenance therapy on simple timed tests of mobility and manual dexterity, and on distance/time parameters of gait was studied in fourteen patients (aged 64 to 88 years). Measurements made 2, 4 and 6 h after morning active and placebo treatments were standardised by taking the pre-treatment measurement on that day as baseline.

3 In a linear model, which allowed for the structure of the study, neither the total time taken by each patient to get up from a chair, walk an individually set distance, turn, return to and sit in the chair, nor the rate of progress at fastening the same set of buttons, was sensitive to the treatment effect.

4 Three of the gait parameters, free walking speed, mean stride length and mean double support time, were sensitive to the treatment effect. Correction for the speed of each walk, caused some reduction in the sensitivity of stride length to treatment effect, but that of double support time remained. Speed, and double support time or stride length, appeared to be complementary in defining the treatment effect.

5 The linear modelling revealed the complexity of the treatment effect. Although active treatment, by comparison with placebo, increased free walking speed ($P = 0.019$), the more levodopa found in the plasma following treatment, ($P = 0.0005$) and the greater the increment in the concentration of its peripheral metabolite, 3-*O*-methyldopa ($P = 0.006$), the less the beneficial effect. This model may reflect reduced uptake into the brain and/or an adverse effect of parent drug or a metabolite.

6 The specificity of free walking speed for the treatment effect was good, as was that of mean stride length, after it had been corrected for speed of each walk, and of mean double support time, after correction for speed and incorporation of the change in lying blood pressure accompanying treatment into the model.

7 The measurements of gait parameters were ranked according to reliability. If, as here, the walk were carried out six times following each treatment and the gait traces measured by hand, free walking speed would be of moderate reliability, but mean stride length and mean double support time of poor reliability. However, the signals generated by the gait analysis apparatus can now be measured directly by computer: computerised gait analysis may prove of adequate reliability, even in frail, elderly patients, for between subject comparisons, as well as being sensitive to and specific for treatment effect within subject.

Correspondence: Dr R. J. Dobbs, Therapeutics in the Elderly, Research Group, CRC, Northwick Park Hospital, Watford Road, Harrow, Middlesex HA1 3UJ

Keywords objective measures brady/hypokinesia gait analysis elderly levodopa/carbidopa

Introduction

Many elderly Parkinsonian patients do not exhibit overt fluctuations in motor state in temporal relation to administration of medication. In these, a subjective assessment is, of course, unlikely to reveal the effects of individual maintenance doses, and the use of objective tests of performance becomes obligatory. In general, the predominant aim in elderly sufferers is to treat their brady- and hypokinesia, and so the question arises as to what measures of brady/hypokinesia to use as outcome criteria. We have investigated a series of measurements of motor performance, comparing their sensitivity to and specificity for the treatment effect, and their reliability. These include the two basic distance/time parameters of gait, stride length and double support time, examined previously (Bowes *et al.*, 1990), along with other, derived, gait parameters. However, even the elegant method of gait analysis selected (Klenerman *et al.*, 1988) requires special equipment and some technical skill, although not the special walkways or clothing, or the complex data extraction of most other methods. Is use of gait analysis justified by the quality of the data it yields, as compared with that of data obtained by simple timed tests of mobility or manual dexterity?

Methods

Patients and study design

The placebo-controlled, balanced, randomised cross-over study of the effect of omission of a morning dose of active treatment, one tablet of the levodopa (100 mg)/carbidopa (25 mg) combination, Sinemet Plus (Merck, Sharp and Dohme Ltd), on measures of brady/hypokinesia in 14 patients (aged 64 to 88 years) with clinical idiopathic Parkinson's disease has been described (Bowes *et al.*, 1990). There, we reported only on determinants of stride length and double support time.

The patients were receiving maintenance therapy with Sinemet Plus alone. Treatments (active and placebo morning doses) were repeated, with gait analysis and timed tests of mobility and manual dexterity being carried out on all 4 days. Measurements of performance made 2, 4 and 6 h after the morning treatment were standardised by taking the pre-treatment measurement on the corresponding day as baseline.

Treatments were repeated a second time, with measurement of mean arterial pressure, supine and erect, and blood sampling for assay of levodopa and a metabolite, 3-*O*-methyldopa (3*O*MD), being carried out in the same relation to the treatments as were those of performance. (All samples from a given patient were included in the same assay (Bowes *et al.*, 1990).) It was assumed that steady state conditions were obtained, so that the fluctuations in plasma concentration on the days when samples were taken were representative of those on the corresponding 'performance testing' days. These procedures were carried out on separate occasions from the tests of brady/hypokinesia in order to avoid the effects of fatigue on performance in a group of subjects, the majority of whom were frail.

A light, low protein breakfast was given 1.5 h before a 10.00 h treatment, no other food or drink being allowed for a similar period of time after. No routine doses of Sinemet Plus were given after the 22.00 h dose on the night before, until 16.00 h on a treatment day, placebo tablets being substituted where appropriate.

Measurements of performance

Patients were subjected to one formal practice of each of the tests described below prior to entry into the study.

Distance/time parameters of gait were measured, using the gait assessment trolley (Klenerman *et al.*, 1988) shown in Figure 1, over a walk of maximum distance 6 m. Gait on a given occasion is represented by two plots, one of distance against time, upward and downward deflections being proportional to distances moved by left and right foot respectively, and the other showing forward velocity of each foot against time (Figure 2). The following were derived from the distance/time trace: each stride length and double support time, free walking speed over the whole walk and cadence (total number of strides made/time taken). The peak velocity achieved by a foot during each swing was obtained from the velocity/time plot.

The mobility test consisted of measuring the time taken by each patient to walk an individually set distance. The distance was constant for, and within the known capabilities of a given patient. The time included that taken to get up from a standard chair, walk the set distance, turn and then return to, and transfer back into the chair.

A buttoning task was selected for the manual dexterity test because of its relevance to activities of daily living. A piece of canvas, 500 by 200 cm, on which a series of nine buttons (graded in size from 3.6 cm down to 0.8 cm diameter, with a 4 cm gap between each) was sewn, and another piece of the same size, with a corresponding series of buttonholes, were used. Each patient was asked to do up the buttons, working from the largest, and easiest, to the smallest. The time taken from the command to start to the fastening of the first button, and that from completion of one fastening to completion of the next was recorded. A maximum time of 2 min was allowed for the test. Subjects might sometimes fail to complete even one fastening. The buttoning time on such occasions was to be regarded as infinite, and the inverse of the buttoning time, that is the buttoning rate, was thus taken as the outcome criterion. Those unable to complete one buttoning would score zero on that occasion.

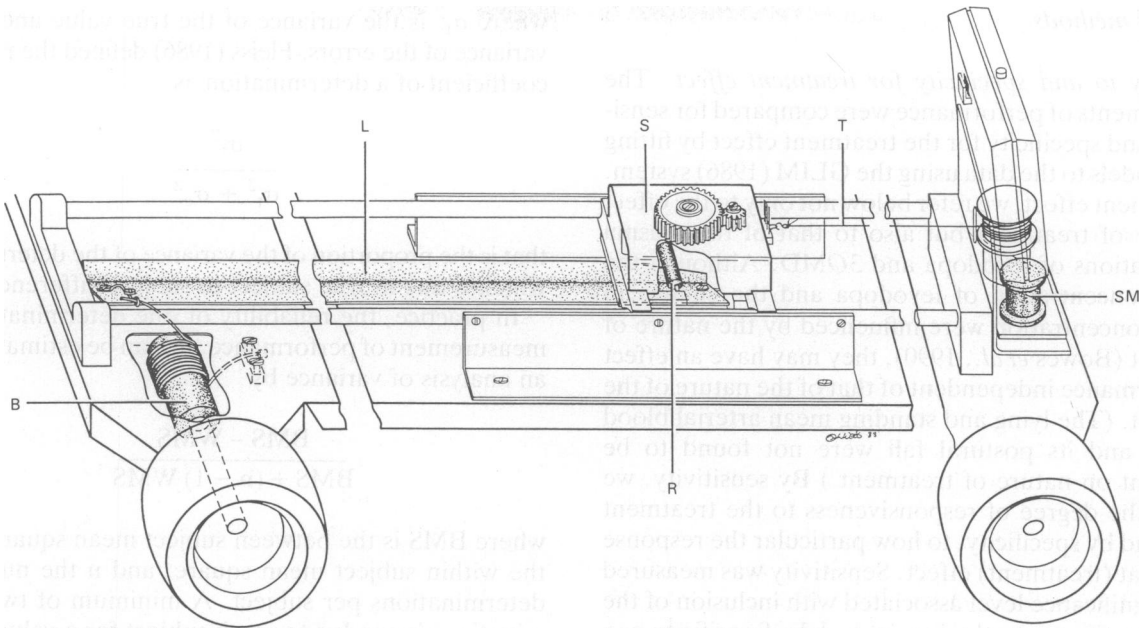

Figure 1 Line drawing of gait assessment trolley. A slide (S) supports a shaft encoder, roller (R) and geared pinch wheels. The slide is normally restrained by thread (T), and stationed by the spring motor (SM). A 3 m length of cotton, clipped to the heels of the patient's shoes, passes between the geared pinch wheels. Walking simultaneously transfers a length of cotton from behind one foot to behind the other, rotates the geared pinch wheels (the larger of which rotates the shaft encoder), and moves the slide forward. As the slide traverses the length of the trolley, roller (R) engages with lever (L) which releases the brake (B), and the trolley rolls forward. If the trolley accelerates too much, the roller disengages and the brake is reapplied. This simple servo-mechanism maintains tension in the cotton. The shaft encoder measures the length of cord transferred, the direction of its rotation denoting which foot has moved. A battery powered, infra-red transmitter sends this encoded information to a receiver connected to a chart recorder.

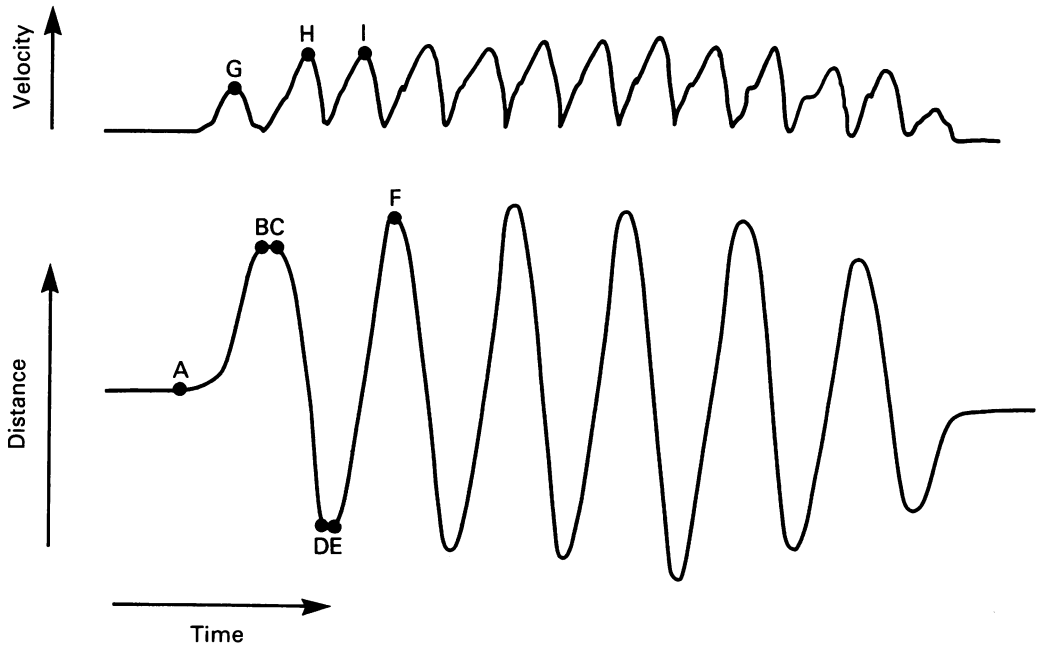

Figure 2 Trace representing the gait of a patient with mild Parkinsonism. In the distance/time plot, points A, C and E correspond to the left foot, the right and left again being lifted off the ground and B, D and F, to the left, right and left again striking the ground. In the velocity/time plot, peak velocity is attained at point G in the first step, and at H and I during the first two swings.

Statistical methods

Sensitivity to and specificity for treatment effect The measurements of performance were compared for sensitivity to and specificity for the treatment effect by fitting linear models to the data using the GLIM (1986) system. By treatment effect, we refer below not only to the effect of nature of treatment but also to that of the plasma concentrations of levodopa and 3*O*MD. Although the plasma concentration of levodopa and the change in 3*O*MD concentration were influenced by the nature of treatment (Bowes *et al.*, 1990), they may have an effect on performance independent of that of the nature of the treatment. (The lying and standing mean arterial blood pressure and its postural fall were not found to be dependent on nature of treatment.) By sensitivity, we refer to the degree of responsiveness to the treatment effect, and by specificity, to how particular the response was to that (treatment) effect. Sensitivity was measured by the significance level associated with inclusion of the treatment effect into the linear models. Specificity was assessed by comparison between the level of significance associated with incorporation of the treatment effect and those associated with incorporation of alternative independent variables.

The formula for the base model, containing the structure of the study together with the grand mean was, in linear model notation (McCullagh & Nelder, 1983), 1 + S + O*T*E, where 1 refers to the grand mean, S to the subjects, O to the order of treatments (active then placebo, or placebo then active), T to the time (2, 4 or 6 h) after the baseline, when the observation was made and E to the exposure to the treatment (initial or repeat). S is regarded as a blocking factor, O a between subject factor, and T and E within subject factors. The significance of extra variables added to the model was assessed as previously described (Bowes *et al.*, 1990). The candidate variable showing the greatest significance was incorporated into the model, irrespective of whether or not it was treatment-related. The procedure was repeated until no more variables approached the required level of significance ($P = 0.01$) for incorporation into the model.

Models were fitted for the following dependent variables: (i) change in mean stride length following treatment, (ii) change in mean double support time, (iii) change in mean peak velocity, (iv) change in free walking speed, (v) change in cadence, (vi) change in time taken for the mobility test, and (vii) change in the rate of performance of the manual dexterity test. Since an effect of the treatment could be to make gait more regular, indices of the variability between strides were also considered as dependent variables: (viii) change in sd of stride length, (ix) change in sd of double support time, and (x) change in sd of peak velocity.

Reliability of measurements of performance When a determination of a measurement is made, the observed value, x, can, in theory, be broken down into two parts: the underlying true value, t, and an error of determination, e. Then, the variance of the observed value, σ^2, is given by

$$\sigma_t^2 + \sigma_e^2$$

where σ_t^2 is the variance of the true value and σ_e^2 the variance of the errors. Fleiss (1986) defined the reliability coefficient of a determination as

$$\frac{\sigma_t^2}{\sigma_t^2 + \sigma_e^2}$$

that is the proportion of the variance of the determination which is due to true subject to subject differences.

In practice, the reliability of one determination of a measurement of performance, R, can be estimated from an analysis of variance by

$$\frac{\text{BMS} - \text{WMS}}{\text{BMS} + (n-1)\,\text{WMS}},$$

where BMS is the between subject mean square, WMS the within subject mean square, and n the number of determinations per subject. A minimum of two determinations is needed for each subject for a value of R to be calculated. If the measurement is reliable, BMS would be large compared with WMS, so R would be large. The maximum value for R is 1: this is achieved when the determination is exact. Fleiss (1986) rated R values less than 0.4 as signifying poor reliability, greater than 0.4 but less than 0.75 as fair to good, and greater than 0.75 as excellent. There is no minimum for R since it is possible for WMS to be greater than BMS: R will then assume negative values. Small positive and negative values for R indicate that there is a lot of error in each determination, that is a single determination would be an unreliable estimate of the true value.

Reliability can be increased by taking several determinations under the same conditions and using the mean of those determinations. The reliability of the mean of n determinations, R_n, is estimated from R by:

$$R_n = \frac{n.R}{1 + (n-1).R},$$

where R is the reliability of a single determination (as above). Alternatively, the number of determinations, n_s, required to give a reliability for the mean of those determinations of a specified value, R_s, can be calculated:

$$n_s = \frac{R_s(1 - R)}{R\,(1 - R_s)}$$

If a series of determinations are made under different conditions in each subject, then the reliability coefficient can be adjusted for this stratification (Fleiss, 1986). (This is also called adjusting for anchor points (Winer, 1971).) Omitting to adjust for stratification would result in an artificially low value for reliability. Here, single determinations of each measurement of performance were made 2, 4 and 6 h post-medication, repeated for both treatments, and the changes from baseline calculated. The values of reliability for these within day changes have been adjusted for the timing of the observation after the baseline and the exposure to the treatment. Reliability coefficients were compared for the dependent variables (i) to (x) described above.

Results

Sensitivity to and specificity for treatment effect

The data on the measurements of performance, standardised according to the corresponding pre-treatment values, are summarized in Figures 3 and 4. The change in performance, expressed as the post-treatment value minus the pre-treatment value, gave homogeneous and normally distributed residuals in the case of mean stride length, peak velocity and free walking speed. For mean double support time, cadence, the mobility test time, buttoning rate, and the s.d.s of stride length, double support time and peak velocity, it was necessary to express the dependent variable as the log of ratio of post-treatment to pre-treatment value.

The F ratios, produced when each of a series of variables was added separately to the base model for each standardised measurement of performance, are shown in Table 1. (For each dependent variable, there

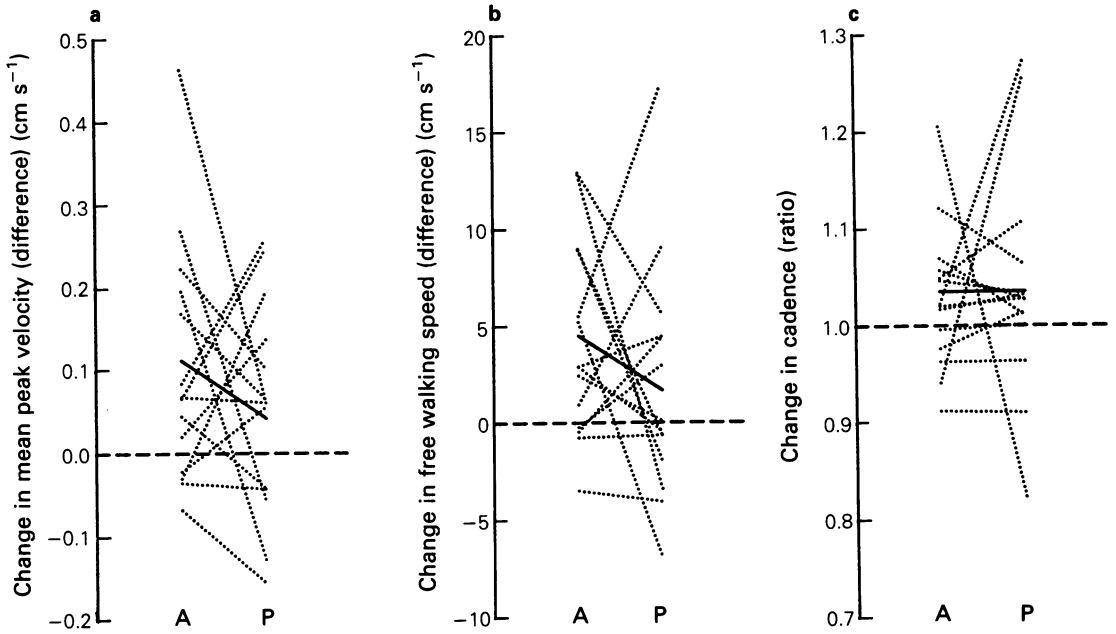

Figure 3 Effect of active (A) and placebo (P) treatments on gait, in individual patients (.....) and averaged over the group (——). Data on a) mean peak velocity, b) speed and c) cadence are shown. (Data on mean stride length and mean double support time have previously been summarised (Bowes *et al.*, 1990).) For mean peak velocity and free walking speed, the average value of the difference, post- minus pre-treatment value, for the three post-treatment sessions on the two exposures to a treatment in each patient, is shown. For cadence, the average value of the ratio of the post- to pre-treatment value is given. Of the possible 84 values, (i.e. six per each of 14 patients) on each treatment, the missing data were as follows: for a) and c), 8 on A and 7 on P; for b) 8 on each A and P.

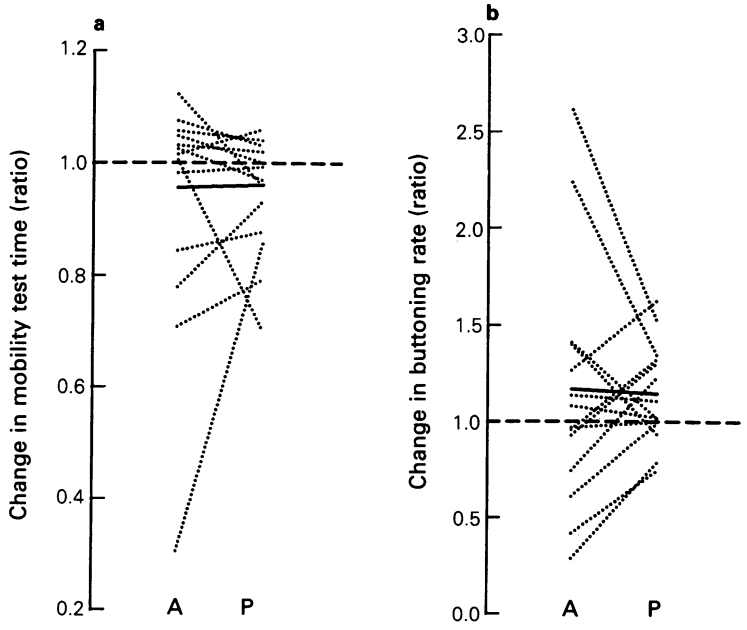

Figure 4 Effect of active (A) and placebo (P) treatments on simple timed performance tests, in individual patients (.....) and averaged over the group (——). Data on a) the timed mobility test and b) the buttoning rate are shown. For both, the average value of the ratio of the post- to pre-treatment value, for the three post-treatment sessions on the two exposures to a treatment in each patient, is shown. Of the possible 84 values on each treatment, the missing data were as follows: for a) 15 on A and on P; for b) 3 on each A and P.

Table 1 Results of model fitting for measurements of performance after taking account of structure of study

				F *ratio*‡						
Independent variables	*Mean stride length*†	*Mean double support*†	*Mean peak velocity*	*Speed*	*Cadence*	*Mobility test time*	*Manual dexterity test rate*	*s.d. stride length*	*s.d. double support time*	*s.d. peak velocity*
Nature of treatment	8.49*	12.06**	3.01	5.67	0.05	0.43	0.02	3.54	4.46	0.29
Pre-treatment speed	16.10***	0.42	0.74	—	6.20	0.59	3.80	0.95	0.04	1.28
Plasma concentration										
Post-treatment levodopa	3.23	12.33**	0.74	0.27	0.20	1.61	0.04	3.15	3.68	1.05
Post-treatment 3*O*MD	0.05	2.59	0.10	0.00	4.17	3.63	3.62	0.20	3.00	1.54
Pre-treatment 3*O*MD	0.02	4.66	0.58	0.47	0.36	3.58	0.36	0.55	0.20	0.00
Change in 3*O*MD	0.00	0.24	0.34	0.64	1.95	0.20	1.34	0.18	1.59	1.68
Mean arterial blood pressure										
Post-treatment lying	0.26	14.55**	2.92	0.70	5.72	1.81	2.14	0.23	1.46	0.79
Pre-treatment lying	5.46	9.71*	4.12	2.94	5.17	3.83	0.35	1.04	13.43**	0.01
Change lying	5.46	33.92***	0.25	4.19	15.30**	7.49*	1.86	2.21	12.59**	0.67
Post-treatment standing	1.17	30.40***	0.00	2.29	12.27**	14.22**	1.25	5.60	14.80**	1.22
Pre-treatment standing	4.78	3.79	6.17	0.18	10.44*	1.58	0.41	0.16	0.01	0.42
Change standing	0.06	15.05**	3.00	0.16	16.43***	9.10*	0.91	0.77	2.62	1.10
Post-treatment postural fall	3.37	0.08	6.35	0.70	0.00	4.95	0.07	1.20	4.79	0.00
Pre-treatment postural fall	6.60	0.52	15.64***	2.21	0.35	0.35	0.10	1.81	7.43*	0.31
Change in postural fall	3.80	0.59	8.10*	1.69	0.43	0.08	0.21	1.43	2.39	0.30

†Reproduced from Bowes *et al.* (1990) for comparison.
‡Assessing significance of independent variables when fitted separately to the base model. (Degrees of freedom ranged from 1,93, to 1,137, for calculation see Bowes *et al.* (1990).
*$P < 0.01$, **$P < 0.001$, ***$P < 0.0001$.

was a maximum of 168 observations, the number of missing observations for each measurement of performance being shown under Figures 3 and 4.) The sensitivity and specificity for the treatment effect revealed by this and further fitting of sequential models, are described below for each of the dependent variables, (i)–(x). (It should be noted that the significance associated with incorporation into these models of the selected variables tended to be much greater ($P << 0.01$) than the critical level suggested in **Methods**.)

(i) *Mean stride length* Mean stride length was sensitive to the nature of treatment ($P = 0.004$), the sensitivity remaining ($P = 0.002$) after allowing for the effect of pre-treatment free walking speed ($P = 0.0001$). Before allowing for pre-treatment speed, the specificity for the treatment effect was relatively good. After allowing for it, specificity was lost: pre-treatment postural fall in mean arterial pressure was nearly as significant ($P = 0.003$) as nature of treatment.

(ii) *Mean double support time* Similarly, this was sensitive to but not specific for the treatment effect. Although both the nature of treatment and the post-treatment levodopa concentration added significantly ($P = 0.0007$ and 0.0006, respectively) to the base model, they were not nearly as significant in this context as were the post-treatment standing, or change in lying, mean arterial blood pressure ($P < 0.0001$ in each case). Moreover, after inclusion of blood pressure, no other variable added significantly to the model.

(iii) *Mean peak velocity* Use of this gait parameter, extracted from the velocity/time plot, appeared to contribute little in terms of sensitivity or specificity. The treatment-related variable showing most effect on mean peak velocity was nature of treatment ($P = 0.085$). This compared with $P = 0.0001$ for a variable which was not treatment-related, namely the pre-treatment postural fall in blood pressure. After adjustment for the postural fall, no significant treatment effect was revealed and no other variable added significantly to the model.

(iv) *Free walking speed* Average speed over the whole walk, as derived from the distance/time plot, was not only sensitive to the treatment effect, but also specific for that effect. Nature of treatment ($P = 0.019$) was the only variable considered which approached the level required for incorporation into the base model. (The next best variable was change in lying blood pressure ($P = 0.043$).) However, after allowing for the effect of nature of treatment, two other treatment-related variables, the post-treatment levodopa and the change in 3*O*MD concentrations, each had highly significant effects on speed ($P = 0.0005$ and 0.0007, respectively). Moreover, after inclusion of nature of treatment and the levodopa concentration, the change in 3*O*MD concentration still retained a significant effect ($P = 0.006$). No other variable added significantly to the model containing these three extra variables.

The overall mean for pre-treatment free walking speed was 39.8 (range 2.6 to 106.6) cm s^{-1} and the mean change was 4.5 (95% C.I. 2.9 to 6.1) cm s^{-1} on active treatment and 1.6 (95% C.I. to 0.0 to 3.3) cm s^{-1} on placebo. That is the increment in speed on active treatment was only 11 per cent, on

average, and there was a small (4 per cent) increment on placebo. The treatment effect was modulated by the post-treatment levodopa concentration and the change in the 3*O*MD concentration. An increment in the levodopa concentration and in the change in 3*O*MD concentration following the dose appeared to reduce the beneficial effect of treatment on speed. On average, for every 100 ng ml^{-1} increase in mean post-treatment levodopa concentration, there would be a decrease in speed of 3.2 cm s^{-1}, and for every 1 $\mu g\ ml^{-1}$ increment in the change in 3*O*MD concentration following treatment, there would be a decrease in speed of 5.9 cm s^{-1}.

(v) *Cadence* No treatment related variable made a significant improvement over the base model (e.g. $P = 0.825$ for nature of treatment). Pre-treatment speed was significant at $P = 0.014$. However, mean arterial pressure could have more importance in this context: in descending order of significance were change in lying ($P = 0.0002$), post-treatment standing ($P = 0.0007$) and pre-treatment standing pressure ($P = 0.0016$). After allowing for the effects of the most significant of these, change in lying pressure, no other variable had a significant effect.

(vi) *Timed mobility test* Some of the measures of mean arterial pressure had a significant effect when incorporated separately into the base model (post-treatment standing ($P = 0.0003$), change in standing ($P = 0.0032$) and change in lying ($P = 0.0073$)), but no treatment related variable was significant (e.g. $P = 0.51$ for nature of treatment). After allowing for post-treatment standing pressure, no other variable had a significant effect.

(vii) *Timed manual dexterity test* This independent variable seemed to be relatively uninfluenced by the extra variables: the smallest probability for a variable added into the base model was for pre-treatment speed ($P = 0.053$). Nature of treatment gave $P = 0.87$ in this context.

(viii) *Standard deviation of stride length* No treatment-related variable had any significant effect when added to the base model (e.g. $P = 0.062$ for nature of treatment). Post-treatment standing pressure ($P = 0.020$) was the only variable considered which approached the level of significance required for incorporation. No other variable improved the model which resulted from incorporating post-treatment standing pressure into the base model.

(ix) *Standard deviation of double support time* As with mean double support time itself, several measures of mean arterial pressure were found to be significant when added to the base model: in descending order, these were post-treatment standing ($P = 0.0002$), pre-treatment lying ($P = 0.0004$), change lying ($P = 0.0005$) and pre-treatment postural fall ($P = 0.007$). The treatment-related variable having most effect was nature of treatment ($P = 0.037$). After adjusting for post-treatment standing pressure, pre-treatment postural fall had a significant effect ($P = 0.006$), but once the latter variable was incorporated, no other variable contributed significantly to the model.

(x) *Standard deviation of peak velocity* No variable significantly added to the base model. Nature of treatment gave $P = 0.6$ in this context.

Value of gait parameters after correction for speed of each walk

The mean stride length, mean double support time and free walking speed all appeared to be sensitive to the treatment effect, but was the sensitivity of the former two gait parameters simply a reflection of the influence of speed on them? The variable post-treatment speed was thus added to the series of candidate variables, and models refitted for mean stride length and mean double support time. The post-treatment speed had a much more significant ($F = 308.00$, $P \ll 0.0001$) effect on mean stride length, than did pre-treatment speed ($F = 16.10$, $P < 0.0001$). Interestingly, after allowing for the effect of post-treatment speed, instead of the pre-treatment, the previous sensitivity to the nature of treatment was lost ($F = 2.69$, $P = 0.1$). Only the variables post-treatment levodopa concentration ($F = 6.51$, $P = 0.012$) and change in 3*O*MD concentration ($F = 6.44$, $P = 0.013$), approached the level of significance for incorporation into the model. However, after incorporation of post-treatment levodopa concentration, no other variable significantly improved the model.

The post-treatment speed was also the most significant ($F = 65.81$, $P < 0.0001$) of the candidate variables, when they were added individually into the base model for mean double support time. After including post-treatment speed, the change in lying blood pressure had the most significant effect ($F = 20.57$, $P < 0.0001$). After incorporation of change in lying pressure and post-treatment speed, post-treatment levodopa concentration contributed most significantly ($F = 13.88$, $P = 0.0003$) to the model, and, after incorporation of the levodopa concentration also, inclusion of the change in 3*O*MD concentration had the most significant effect ($F = 8.41$, $P = 0.004$). No other variable reached the required level of significance for incorporation into the resultant model for mean double support time. This model, notably, did not contain the nature of treatment. A beneficial effect on double support time (that is a shortening) was seen with increasing post-treatment levodopa concentrations. The effect was modulated by the change in 3*O*MD concentration following the dose, an increment in the change in 3*O*MD reducing the benefit.

Timing of post-treatment observations

Analysis of the final models showed that there was no significant effect of timing 2, 4 or 6 h after baseline (or order of, or exposure to, the treatments), on any of the ten dependent variables. Moreover, there were no significant interactions between timing and the order of, or exposure to, the treatments.

Overall effect of mean arterial blood pressure on measurements of performance

Attention has already been drawn to the significance of the blood pressure variables when added individually to the base model for each measure of performance (Table 1). Essentially, the tendency was that the lower the post-treatment lying and standing pressure, and the smaller the increment, or greater the decrement, in these pressures from their pre-treatment values, the more hurried the performance. The hurried gait was manifest in a decreased and less variable double support time and an increase in cadence, whilst the mobility test was performed faster. A trend towards a negative change in pressure (post- minus pre-treatment) can, of course, be achieved, not only by a decrease in the post-treatment pressure, but also by an increase in the pre-treatment pressure, or both. An increase in the pre-treatment standing and lying pressures was, indeed, associated with a hurried gait, again as evidenced by both an increase in cadence and a decrease in the mean and s.d. of the double support time.

The incorporation of pre-treatment postural fall in blood pressure (lying minus standing) into a model for mean stride length, containing baseline speed, was significant: the greater the postural fall, the greater the increment in stride length following treatment. Similarly, the incorporation of pre-treatment postural fall into the base models for mean peak velocity and s.d. of double support time was significant: the greater the pre-treatment postural fall, the greater the peak velocity attained during the strides and the more regular the gait with respect to double support time. This was probably explained by an acquired tolerance to the fall as the day progressed. An increment in the change in postural fall (post- minus pre-treatment) was detrimental to the peak velocity attained: presumably the patients were not able to compensate when the postural fall actually increased.

Reliability

Table 2 gives the calculated reliability for a single determination of each of the ten dependent variables, and for the mean of the six determinations, as in the present protocol. An estimate of the number of determinations required to produce moderate reliability ($R = 0.5$) and good reliability ($R = 0.75$) is shown. It appears that the simple timed performance tests were more reliable than gait anlaysis. Indeed, in the case of the timed mobility test, the reliability for the mean of six determinations was good. Of the gait parameters, speed and the mean and standard deviation for peak velocity fell within Fleiss's definition of fair to good reliability (see Statistical methods), when six determinations were available. However, cadence, mean double support time and mean stride length would require repetition of the walk eleven, eighteen and nineteen times respectively, to produce moderate reliability.

Discussion

In using the terms specificity and sensitivity, we have taken the analogy with a drug assay, where sensitivity refers to the minimum change in a drug concentration which can be detected and specificity to the extent to which other substances interfere with the measurement. The present study constitutes a bioassay: the sensitivity of a given outcome criterion to the treatment effect is measured, together with its specificity for that effect. As in many bioassays, there is no 'gold standard' for quality control of reproducibility. Instead, we have estimated the reliability coefficient for each outcome criterion.

In patients with idiopathic Parkinson's disease, who do exhibit overt fluctuations in motor performance in temporal relation to individual doses of medication, all that may be required, in a randomised crossover study, to distinguish between treatments is assessment of whether a patient is 'on' or 'off', and of the proportion of time spent in either state (Colman *et al.*, 1989). Patients who have received maintenance levodopa therapy for two or more years, and in whom the onset of Parkinson's disease was early, are particularly vulnerable to sudden switches in motor performance. However, in such patients, application of the subjective motor examination from the Unified Parkinson's Disease Rating Scale (Fahn & Elton, 1987) has shown response to be graded (Colman *et al.*, 1989). Moreover, performance in the 'on' state may be impaired by drug-induced dyskinesia and dystonia. Thus, even in those with the 'on-off' syndrome, there is a case for objective assessment of

Table 2 Comparative reliability and requirement for replication of a series of measurements of performance

Measurement	*Reliability of n determinations* n = 1	n = 6	*Number of determinations required for reliability of* 0.50	0.75
Mean stride length	0.05	0.24	19.0	56.9
Mean double support time	0.05	0.25	17.7	53.1
Mean peak velocity	0.13	0.48	6.6	19.7
Speed	0.16	0.53	5.3	15.9
Cadence	0.09	0.36	10.8	32.3
Time taken for mobility test	0.38	0.78	1.7	5.0
Time taken for manual dexterity test	0.17	0.56	4.8	14.4
s.d. stride length	0.08	0.35	11.3	33.8
s.d. double support time	-0.04*	—	—	—
s.d. peak velocity	0.14	0.49	6.3	18.8

*N.B. R was negative, the determination was extremely unreliable.

nett benefit instead of simultaneous subjective assessments of the signs of Parkinsonism and the complications of therapy.

In contrast to the above, the group of patients studied here, all elderly and several of them frail, were characterised by having no overt fluctuations in motor performance in relation to administration of a levodopa/decarboxylase inhibitor combination. However, they did have marked variation in performance between days. Thus, they were a challenging group in which to test the sensitivity of any measurement of performance to the effect of medication, and the specificity of the measurement for that effect. Nevertheless, a method of gait analysis, designed for use in clinic or ward, showed potential in defining the effect of omission of a single dose of maintenance therapy in such patients.

The sensitivity and specificity of free walking speed for the treatment effect was good, as was that of mean double support time, after it had been adjusted for speed of each walk and the change in lying blood pressure accompanying treatment. After adjustment of mean stride length for speed of each walk, some of the sensitivity to, but none of the specificity for, the treatment effect was lost: the post-treatment plasma levodopa concentration and the change in that of 3*O*MD were the only other variables which approached the level of significance necessary for incorporation into the sequential linear model. It appears that free walking speed and mean double support time, and possibly also mean stride length, are complementary in defining the treatment effect in patients such as ours. The most discriminant gait parameters may, of course, vary according to the patient subgroup and even the nature of the drug used. Knutsson & Martensson (1971), using interrupted light photography, found that maximal and free walking speeds, swing length and swing time discriminated between sessions of observation conducted pre- and post-introduction of maintenance levodopa therapy: double support time was not discriminant. However, this was an open study of introduction of therapy, and should be interpreted in the light of our finding, in a randomised, cross-over study, of a positive response to a dose of placebo which interrupted maintenance therapy. Gopinathan *et al.* (1981), using a mat with sensors, were able to demonstrate consistent differences between lisuride and placebo with respect to free walking speed, but, surprisingly, to neither stride length nor cadence.

In the work presented, where the gait traces were measured by hand, the six walks performed would have produced data of moderate reliability for a between-subject comparison using free walking speed, but of inadequate reliability for one using mean double support time or mean stride length. Reliability could be improved by increasing the number of observations made under the same conditions. The three determinations of gait parameters made here, between 2 h and 6 h after the treatment, were not significantly different. However, it is obviously necessary in the frail not to increase the frequency of the walks (even when they are as short as 6 m) sufficiently to produce progressive fatigue. Repeating active and placebo days was considered preferable in our patients. In contrast, in those with 'on-off' syndrome, the transient nature of the effect must be a major determinant of the protocol. The signals generated by the shaft encoder of the gait assessment trolley can now be measured by computer (Weller *et al.*, 1989). The computerised method should, since it reduces considerably the error component, increase the sensitivity of the above gait parameters to the (within subject) treatment effect. It may well prove sufficiently reliable for between subject comparisons, where relevant gait parameters are used as the outcome criteria.

Simple objective tests, requiring no specialised equipment beyond, for example a stop watch, tape measure, peg board, pen and paper, or buttons and button holes, have frequently been advocated for inclusion in the outcome criteria of clinical trials in Parkinson's disease (e.g., Broe & Caird, 1973; Brumlik & Boshes, 1966; Calne *et al.*, 1971; Mawdsley, 1970; Mindham, 1976; Nutt *et al.*, 1984; Walker *et al.*, 1972). However, there is little information on their sensitivity to and specificity for treatment effects, and as to their reliability. Interestingly, in this cohort of patients from a major subgroup on maintenance anti-Parkinsonian therapy, typical tests, a timed mobility test and the rate of progress in a buttoning task, showed no sensitivity ($P > 0.5$ in each case) to the treatment effect. Although, when six replicates were available, these tests were of moderate to good reliability, they were of no use in determining efficacy of drug treatment in those without overt response. Free walking speed was useful in this respect, but it was measured directly from the gait trace (from the time a foot first left the ground to the time of the last foot strike). It was, therefore, not confounded by observer and subject reaction times, factors which may be of particular importance over a short distance and in Parkinsonian patients.

The use of gait analysis allowed demonstration of the complex nature of the treatment effect. For example, the beneficial effect of treatment on free walking speed was less at higher post-treatment plasma levodopa concentrations. Further investigation is needed to determine whether there is, in fact, a bell shaped plasma levodopa concentration/response curve. A fall off in response with increasing dose may seem surprising, since the maximum daily dose of levodopa in this group of patients was only 800 mg, whereas patients with 'on/off' syndrome require up to 2 g per day. It does, however, corroborate our clinical experience in patients with late onset idiopathic Parkinson's disease, many of whom do not exhibit overt fluctuations in performance in relation to administration of medication. Such patients have been underrepresented in the literature. The greater the increment in the concentration of the peripheral levodopa metabolite, 3*O*MD, following treatment, the more pronounced was the attenuation of benefit. The fall off in response may be due to reduced uptake of levodopa into the brain, *per se*, or as a result of competition between levodopa and 3*O*MD for uptake (Wade & Katzman, 1975). Alternatively, the reduced response may be a central or peripheral adverse effect of parent drug or a metabolite. The major confounding influence of mean arterial blood pressure on the treatment effect was also revealed. Both the effect of a relatively low post-treatment mean arterial pressure and a large pre-treatment postural fall tended to simulate a beneficial effect of treatment.

In conclusion, gait analysis has potential in the objective assessment of the effect of levodopa therapy, but further work is required to determine whether a combination of

gait parameters discriminates better for that effect than any single parameter. In patients with an overt treatment effect, selection of the best parameter(s), and examination of the treatment effect and any confounding influences, should prove a relatively straightforward task. However, it cannot be assumed that the nature of the effect and its influences will be the same as in the patients described here.

We wish to thank Mrs C. Doré, Head of the Section of Medical Statistics, CRC, for her advice and support and Mrs J. Gilbert for preparing the manuscript.

References

Bowes, S. G., Clark, P. K., Leeman, A. L., O'Neill, C. J. A., Weller, C., Nicholson, P. W., Deshmukh, A. A., Dobbs, S. M. & Dobbs, R. J. (1990). Determinants of gait in the elderly Parkinsonian on maintenance levodopa-carbidopa therapy. *Br. J. clin. Pharmac.*, **30**, 13–24.

Broe, G. A. & Caird, F. I. (1973). Levodopa for Parkinsonism in elderly and demented patients. *Med. J. Aust.*, **1**, 630–635.

Brumlik, J. & Boshes, B. (1966). The mechanism of bradykinesia in Parkinsonism. *Neurology*, **16**, 337–344.

Calne, D. B., Reid, J. L., Vakil, S. D., Rao, S., Petrie, A., Pallis, C. A., Gawler, J., Thomas, P. K. & Hilson, A. (1971). Idiopathic Parkinsonism treated with an extra cerebral decarboxylase inhibitor in combination with levodopa. *Br. med. J.*, **3**, 729–732.

Coleman, R. J., Temlets, J. A., Quinn, N. P., Stahl, S. M. & Marsden, C. D. (1989). A strategy for measuring the therapeutic properties of new anti-Parkinsonian drugs, such as +PHNO, in patients with on-off fluctuations. *Clin. Neuropharmacology*, **12**, 37–45.

Fahn, S. & Elton, R. L. (1987). Unified Parkinson's Disease Rating Scale. In *Recent developments in Parkinson's disease*, Vol. 2, eds Fahn, S., Marsden, C. D., Calne, D. & Goldstein, M., pp. 153–163. Florham Park: Macmillan Healthcare Information.

Fleiss, J. L. (1986). *The design analysis of clinical experiments*. Chapter 1. New York: Wiley.

GLIM (1986). *Release 3.77 Generalised linear interactive modelling*. Oxford: Numerical Algorithms Group Ltd.

Gopinathan, G., Teravainen, H., Dambrosia, J. M., Ward, C. D., Sanes, J. N., Stuart, W. K., Evarts, E. V. & Calne, D. B. (1981). Lisuride in Parkinsonism. *Neurology*, **31**, 371–376.

Klenerman, L., Dobbs, R. J., Weller, C., Leeman, A. L. & Nicholson, P. W. (1988). Bringing gait analysis out of the laboratory and into the clinic. *Age Ageing*, **17**, 397–400.

Knutsson, E. & Martensson, A. (1970). Quantitative effects of L-dopa on different types of movements and muscle tone in Parkinsonian patients. *Scand. J. Rehabil. Med.*, **3**, 121–130.

Leeman, A. L., Hughes, J., Bowes, S. G., O'Neill, C. J. A., Weller, C., Clark, P., Deshmukh, A. A., Nicholson, P. W., Dobbs, S. M. & Dobbs, R. J. (1989). Role of gait analysis in the assessment of Parkinsonism in old age. In *Current problems in Neurology*, **9**, Neural Mechanisms in Disorders of Movement, eds Crossman, A. R. & Sambrook, M. A., pp 319–323.

Mawdsley, C. (1970). Treatment of Parkinsonism with levodopa. *Br. med. J.*, **1**, 331–337.

McCullagh, P. P. & Nelder, J. A. (1983). *Generalised linear models*. London: Chapman and Hall.

Mindham, R. H. S. (1976). Assessment of drug-induced extrapyramidal reactions and of drugs given for their control. *Br. J. clin. Pharmac.*, **3**, Suppl. 2, 395–400.

Nutt, J. G., Woodward, W. R., Hammerstad, J. P., Carter, J. H. & Anderson, J. L. (1984). The 'on-off' phenomenon in Parkinson's disease: relation to levodopa absorption and transport. *New Engl. J. Med.*, **310**, 483–488.

Wade, L. A. & Katzman, R. (1975). 3-O-Methyldopa and inhibition of L-dopa at the blood–brain barrier. *Life Sci.*, **17**, 131–136.

Walker, J. E., Albers, J. W., Tourtellotte, W. W., Henderson, W. G., Potvin, A. R. & Smith, A. (1972). A qualitative and quantitative evaluation of amantadine in the treatment of Parkinson's disease. *J. chronic Dis.*, **25**, 149–182.

Weller, C., Dobbs, R. J., Tate, M. A., Telford, A. & Klenerman, L. (1989). In *Biotelemetry* X. Proceedings of tenth international symposium on biotelemetry, ed. Amlaner, C. J., pp 575–580. Fayetteville: The University of Arkansas Press.

Winer, B. J. (1971). *Statistical principles in experimental design*, Chapters 4 and 5. New York: Wiley.

(Received 14 June 1990, accepted 24 September 1990)