according to their DST results: patients with an abnormal DST had a significantly worse score (compared to those with a normal DST) on both tests.

The level of depression was assessed on the Krawiecka Scale. There was no significant difference among the three groups of patients.

Discussion

We have found that patients with Type II schizophrenia have an abnormal DST, compared to Type I schizophrenia and to normal controls. The mixed-type group occupies an intermediate position between Type I and Type II. We also found that Type II tended to have a lower level of cognitive function. There was a strong association between an abnormal cognitive function and an impaired DST.

These findings are in line with the descriptions of Type II schizophrenia by Crow (1980), who summarised the evidence for structural abnormality in the brains of Type II schizophrenics. These results are of particular interest in view of the report of abnormal DST in patients with senile dementia (Coppen *et al*, 1983).

This investigation supports the biological distinction between Type I and Type II schizophrenia. The prognostic significance, extent of reversibility and implications for treatment of an abnormal DST response will be matters for further research.

Acknowledgements

We are grateful to Mrs J. Harwood and Dr. C. Swade for assistance.

References

COPPEN, A., ABOU-SALEH, M., MILLN, P., METCALFE, M., HARWOOD, J. & BAILEY, J. (1983) Dexamethasone suppression test in depression and other psychiatric illness. *British Journal of Psychiatry*, **142**, 498–504.

CROW, T. J. (1980) Molecular pathology of schizophrenia; more than one disease process? *British Medical Journal*, **i**, 66–68.

FEIGHNER, J. P., ROBINS, E., GUZE, S., WOODRUFF, R. A., JR., WINOKUR, G. & MUNOZ, R. (1972) Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, **26**, 57–62.

INGLIS, J. (1959) A paired-associate learning test for use with elderly psychiatric patients. *Journal of Mental Science*, **105**, 440–443.

KRAWIECKA, M., GOLDBERG, D. & VAUGHAN, M. (1977) Standardised psychiatric assessment scale for rating chronic psychotic patients. *Acta Psychiatrica Scandinavica*, **55**, 299–308.

WECHSLER, D. (1945) *Wechsler Adult Intelligence Scale Manual*. New York: Psychology Corporation.

D. Saffer, MB, ChB, DPM, *Clinical Assistant, Department of Psychiatry, Epsom District Hospital, Epsom, Surrey*

M. Metcalfe, DipPsych, *Research Psychologist*

*A. Coppen, MD, DSc, FRCP, FRCPysch, *Director, MRC Neuropsychiatry Research Laboratory, West Park Hospital, Epsom, Surrey KT19 8PB*

*Correspondence

(*Accepted 5 March 1985*)

British Journal of Psychiatry (1985), **147**, 723–726

Measuring Abnormal Movement in Tardive Dyskinesia: A Pilot Study

WILSON R. FIRTH and MARK H. ARDERN

Despite the many techniques available, research into tardive dyskinesia (TD) is handicapped by the difficulty of measuring abnormal movement. Conflicting reports of both prevalence and response to treatment may, in part, be due to this problem. In this pilot study, two methods of measurement were evaluated—video recording and frequency counts. Standardised Video Recording (SVR) has a number of advantages over clinical observation for the research worker: SVRs can be stored and if their order of presentation is randomised, raters are more 'blind'. Conversely, a video rater cannot elicit

https://doi.org/10.1192/bjp.147.6.723 Published online by Cambridge University Press

movements and observes only a two-dimensional picture which may distort or miss certain movements (Gardos & Cole, 1980). Although good agreement between clinical and SVR ratings in diagnosing TD has been reported (Barnes & Trauer, 1982), it has yet to be shown that SVR displays specific movements reliably. We therefore wished to compare assessments made from clinical procedures with those made from video recordings.

Assessment of abnormal movements by counting their frequency has mostly been used to measure changes following drug challenge. The frequency is measured by counting how often a movement occurs in a given time, usually 30 seconds. Unfortunately, TD movements are often irregular, so that frequencies may vary considerably from one count to the next. A reliable estimate of the average frequency therefore requires that only regularly occurring movements are selected for measurement and that they are observed for a sufficiently long time. Studies using this method (Kazamatsuri *et al*, 1972a, b, c, 1973; Davis *et al*, 1975; Roxburgh, 1970; Fann & Lake, 1974; Growdon *et al*, 1977; Moore & Bowers, 1980) have not discussed these problems in detail, and observation of irregular movements led Moore & Bowers to comment that excessive fluctuation of baseline counts made response to drug challenge difficult to interpret. We have therefore developed a method of selecting movements and of finding the minimum time required to assess their frequency.

Method

(1) *Comparison of clinical and videotape ratings of TD movements*

Ten chronic schizophrenic patients gave their informed consent to participate in the study; they had all been treated with neuroleptics for at least two years. All had TD, as diagnosed by two experienced clinicians, and had at least one movement rated 4–6 (moderate–severe) on the Abbreviated Rockland Rating Scale (ARRS, Simpson *et al*, 1979). The ARRS was chosen because it allows study of specific movements. To compare clinical with video ratings, the patients were first rated on the ARRS by the clinical rater (CR), who followed the recommended assessment procedure; immediately afterwards, they were videotaped for one minute. In order to standardise for posture and concentration, each patient was seated and asked to memorise a piece of standard prose, mounted on a music stand in front of him (Barnes & Kidger, 1979; Kidger *et al*, 1980). Six months later, these videorecordings were rated on the ARRS form by two raters, one of whom had been the clinical rater (now designated C-VR), and the other an independent rater (I-VR).

Agreement between clinical ratings and each of the videotape ratings, and between the two videotape ratings was calculated for ten of the 17 ARRS items scored: periocular area, movement of lips, chewing movements, bon-bon sign, tongue protrusion, rocking movements of neck/trunk, torsion movements of neck/trunk, movements of fingers and wrists, movements of ankles and toes, stamping movements.

Seven items were thus excluded: those irrelevant to TD (akathisia), unspecified movements (items described as 'other movement'), or movements impossible to see with this method of videotaping (intraoral tongue movements and axial hyperkinesis of the trunk—observed with patients standing). If any particular movement's presence was scored 'uncertain' by a rater, then the ratings for that movement were discarded from agreement calculations. Agreements between ratings for the remaining ten items were assessed in two ways. Firstly, the percentage of concordant present–absent ratings out of the total number considered was calculated. Secondly, the agreement co-efficient 'Weighted Kappa' (Kw) was calculated by the method of Ciccheti (1976): this compared raters' scores on a severity scale of 1–5, the uncertainty score '2' being excluded from comparisons. A rater's 'total score' for each of the patients was the aggregate of scores for all 16 TD items on the ARRS, akathisia being excluded from the total.

Of the 100 items rated (ten patients × ten items), 23 were scored 'uncertain' by one or more raters, which left 77 to be considered for agreement calculations. 'Uncertainty' ratings were significantly more common in assessments by CR ($n = 16$) by either C-VR ($n = 5$) or I-VR ($n = 4$), (in both cases $P<.05$, chi squared). CR rated intraoral tongue movement present in six cases and axial hyperkinesis of the trunk in two, but these movements were not seen by either C-VR or I-VR.

Present–absent (P-A) agreements for each of the ten ARRS items was >75% between the clinical rater and each of the videotape raters; for the 77 movements, mean P-A agreement was 91%. There was full P-A agreement between C-VR and I-VR. Nevertheless, because of the small numbers involved, conclusions drawn from these results, particularly in the specific items listed, must be limited.

There was good agreement between the raters' scores for these 77 movements, the co-efficients of agreement (Kw) being as follows: for CR and C-VR:—0.82; for CR and I-VR:— 0.79; for C-VR and I-VR:—0.77.

The mean total score by CR (27.6 ± 4) was significantly greater (Wilcoxon test) than that by either C-VR (25.1 ± 2.3, $p<.05$) or I-VR (24.0 ± 2.0, $p<.01$). When intraoral tongue movements and axial hyperkinesis scores were excluded, the mean total by CR (24.6 ± 2.2) was slightly, but not significantly greater than those by C-VR (24.5 ± 1.6) and I-VR (23.6 ± 1.9).

(2) *Selection of TD movements suitable for frequency counts: theoretical basis*

Events which are evenly dispersed in time should conform to a Poisson Distribution (Moroney, 1965). In selecting movements suitable for assessment by this method, we

https://doi.org/10.1192/bjp.147.6.723 Published online by Cambridge University Press

therefore defined regularly occurring movements as those in which the frequency distribution of counts conformed to a Poisson Distribution. Irregular movements occur in bursts with interspersed periods of relative inactivity; they would not fit a Poisson Distribution because this requires that movements are independent of one another, not grouped together. The proposed method sought to distinguish between those movements which conformed to a Poisson Distribution from those which did not.

(3) *Procedure*

Four chronic schizophrenics fulfilled the selection criteria described above. The nine movements studied were all those rated present (≥3 on the ARRS), regardless of their apparent regularity. Videorecordings were made as above, except that recordings went on for 15–22 minutes. The number of times each abnormal movement occurred in 30 seconds was counted for successive 30-second periods throughout the videotape; a count was discarded if visibility was obscured for any part of any period. This gave an estimator sample of 30–44 counts for each movement.

The mean count and standard deviation of each movement was calculated from this sample. Three further calculations were made. Firstly, inter-rater reliability for two raters was calculated from a five-minute sample taken from the mid-period of the videotape (Pearson correlation co-efficient). Secondly, goodness of fit of the count values to a Poisson Distribution was calculated by the chi-squared test (see Appendix I). Thirdly, a calculation was made from the estimator sample of the minimum observation time required to make a reasonably accurate estimate of a movement's mean count. This was termed Minimum Sampling Time (MST) and was calculated from the variability of each movement (Appendix II).

Results

Inter-rater reliability was high for all movements (>.9) except for jaw movements in one subject, for which it was .75.

Mean frequency counts ranged from .9–25.0 movements per 30 seconds; there was no clear relationship between a movement's mean count and its goodness of fit to a Poisson Distribution.

Frequency counts of four movements (Patient no. 2: Left Hand, Left Foot; Patient no. 3: Jaw; Patient no. 4: Eyes) did not differ significantly from a Poisson Distribution ($p>.1$). In each case, Minimum Sampling Time was three minutes or less. The remaining five movements (Patient no. 1: Mouth, Right Foot, Left Foot; Patient no. 2; Right Hand and Right Foot) did not fit a Poisson Distribution ($p<.005$) and their Minimum Sampling Times were respectively, 5, 6, 9, 21 and 8 minutes.

Discussion

The study showed good, but not complete, agreement between clinical and videoraters for both specific movements and total scores. Videoraters picked up fewer movements, even within the ten movements rated, and their total scores were slightly less, but these differences were not statistically significant. The clinical rater had more 'uncertainty' scores. It is possible that the longer time and flexibility available to the clinical rater accounted for these differences. Where assessments from video include ratings of concealed trunk and tongue movements, as in studies which delineate patterns of TD (Barnes *et al*, 1980), additional procedures are required to make these movements visible.

Longer videorecordings, used to study the frequency distribution of counts for each of nine movements, allowed selection of regularly occurring movements by estimating goodness of fit to a Poisson Distribution. We therefore suggest that in studies which require a series of measurements, e.g. drug challenge tests, preliminary videotaping is necessary to establish a movement's regularity and its countability. The duration of further assessments of a movement's mean frequency may need to be longer than the Minimum Sampling Times quoted here, so that significance tests can be applied. Frequency counting is a simple, sensitive, and often reliable method of measuring TD; proper selection of movements suitable for assessment by this method is likely to enhance its value.

Appendix I

Assessing goodness of fit to a Poisson distribution

The mean count for each movement was calculated from the estimator sample. The expected number of counts for each frequency value was calculated from Poisson probability charts, which were tabulated with observed numbers and regrouped so that cell numbers were ≥5. The chi-squared test was applied to test goodness of fit (Moroney, 1965).

Appendix II

Calculation of Minimum Sampling Time (MST)

MST was chosen to express the variability of each movement as the amount of observation time required to be 95% certain that the mean value calculated from this sample is accurate to within the square root of the population mean. The square root was chosen so that MST was not affected by the magnitude of the mean. The mean of the estimator sample is assumed to be that of the population (x). The formula is derived from that of SEm, where the square root of the mean is substituted for SEm as the desired level of accuracy (Moroney, 1965).

$$\sqrt{\bar{x}} \simeq \frac{SD \times 1.96}{\sqrt{n}}$$

$$\therefore n \simeq \frac{SD \times 1.96^{2}}{\sqrt{\bar{x}}}$$

https://doi.org/10.1192/bjp.147.6.723 Published online by Cambridge University Press

This is expressed in minutes, that is n/2. In a Poisson Distribution $\sqrt{\bar{x}} = SD$, so that $n \simeq (1.96)^2 \simeq 4 = 2$ minutes. Results were rounded up to the nearest minute, so for movements which conform to a Poisson Distribution, MST = 3 mins.

Acknowledgements

We are most grateful for assistance to Drs T. Kolakowska, P. Cooper and H. Wilkes, Pat Finlay and Professor M. Gelder. Dr Firth's research attachment was funded by the Wellcome Foundation.

References

BARNES, T. R. E. & KIDGER, T. (1979) Tardive dyskinesia and problems of assessment. In *Current Themes in Psychiatry* (eds. R. Gaind & B. Hudson). Vol. 2. London: Macmillan.

—— —— TRAUER, T. & TAYLOR, P. (1980) Reclassification of the tardive dyskinesia syndrome. In *Long Term Effects of Neuroleptics* (eds. F. Cattabeni *et al*). New York: Raven Press.

—— —— —— (1982) Reliability and validity of a tardive dyskinesia rating technique. *British Journal of Psychiatry*, **140**, 508–515.

CICCHETI, D. V. (1976) Assessing inter-rater reliability for rating scales: resolving some basic issues. *British Journal of Psychiatry*, **129**, 452–456.

DAVIS, K. L., BERGER, P. A. & HOLLISTER, L. E. (1975) Choline for tardive dyskinesia. *New England Journal of Medicine*, **293**, 152.

FANN, W. E. & LAKE, R. (1974) On the co-existence of parkinsonism and tardive dyskinesia. *Diseases of the Nervous System*, **35**, 324–326.

GARDOS, G. & COLE, J. D. (1980) Problems in the assessment of tardive dyskinesia. In *Tardive Dyskinesia: Research and Treatment* (eds. R. C. Smith, W. E. Fann & J. M. Davis). New York: Spectrum Books.

GROWDON, J. H., HIRSCH, M. J., WURTMAN, R. J. & WIENER, W. (1977) Oral choline administration to patients with tardive dyskinesia. *New England Journal of Medicine*, **297**, 524–527.

KATZAMATSURI, H. B., CHIEN, C. P. & COLE, J. O. (1972a) Treatment of tardive dyskinesia. I. Clinical efficiency of a dopamine depleting agent, tetrabenazine. *Archives of General Psychiatry*, **27**, 95–99.

—— —— —— (1972b) Treatment of tardive dyskinesia. II. Short term efficiency of dopamine blocking agents haloperidol and thiopropazate. *Archives of General Psychiatry*, **27**, 100–103.

—— —— —— (1972c) Treatment of tardive dyskinesia III. Clinical efficacy of a dopamine competing agent, methyldopa. *Archives of General Psychiatry*, **27**, 824–827.

KATZAMATSURI, H. B., CHIEN, C. P. & COLE, J. O. (1973) Long term treatment of tardive dyskinesia with haloperidol and tetrabenazine. *American Journal of Psychiatry*, **130**, 479–483.

KIDGER, T., BARNES, T. R. E., TRAUER, T. & TAYLOR, P. J. (1980) Subsyndromes of tardive dyskinesia. *Psychological Medicine*, **10**, 513–520.

MOORE, D. C. & BOWERS, M. B. (1980) Identification of a subgroup of tardive dyskinesia by pharmacologic probes. *American Journal of Psychiatry*, **137**, 1202–1205.

MORONEY, M. J. (1965) Facts from Figures. Harmondsworth: Pelican Books.

ROXBURGH, P. A. (1970) Treatment of persistent phenothiazine induced oral dyskinesia. *British Journal of Psychiatry*, **116**, 277–280.

SIMPSON, G. M., LEE, J. H., ZOUBOK, B. & GARDOS, G. (1979) A rating scale for tardive dyskinesia. *Psychopharmacology*, **64**, 171–179.

*Wilson R. Firth, BSc, MB, ChB, MRCP, MRCPsych, *Senior Registrar, Littlemore Hospital, Oxford*

Mark H. Ardern, MB, ChB, MRCPsych, *Senior Registrar, Warneford Hospital, Oxford*

*Correspondence

(*Accepted 14 August 1984*)

https://doi.org/10.1192/bjp.147.6.723 Published online by Cambridge University Press