British Journal of Psychiatry (1988), **153**, 354–358

Criteria for Evaluating Improvement in Schizophrenia in Psychopharmacological Research (With Special Reference to Gamma Endorphin Fragments)

RAHUL MANCHANDA, STEVEN R. HIRSCH and THOMAS R. E. BARNES

A review of treatment trials with DTγE revealed widely discrepant results. Relevant variables were the variety of measures employed for monitoring psychotic symptoms, and the different criteria used to judge the degree of improvement. The authors suggest a uniform outcome criterion for early trials of new treatments, which would generate more consistent and comparable results between studies, and give a stronger indication of the value of the treatment under test. When the data from the various treatment trials of DTγE were reanalysed, applying a uniform outcome criterion of improvement of a change of 80% or more on rating-scale score, the results were more consistent than would have been suspected from the original reports.

Treatment trials invariably provide data on the differences in rating-scale scores between treatment groups. Such information may not, in itself, allow patients to be classified as improved or not improved. Some indication of the number of patients 'responding' to the treatment is desirable, which requires a judgement as to what is or is not to be regarded as improvement. The basis for such a judgement varies from study to study, and this variation may help to explain the lack of consistency in results from different centres, when reporting on the effectiveness of new treatments.

This issue is dealt with specifically in regard to the discrepancies between the results of different trials of gamma-endorphin fragments, namely des-tyrosine-γ-endorphin (DTγE or LPH 62-77) and des-enkephalin-γ-endorphin (DEγE or LPH66-77) for the treatment of schizophrenia. These gamma endorphins were chosen for the following reasons:

1. They are relatively new and scientifically interesting compounds, and occur naturally in the pituitary of man and the rat (Loeber *et al*, 1979; Verhoef *et al*, 1980); they have very specific behavioural effects in animals, resembling those of antipsychotic drugs (De Wied *et al*, 1978*a*; Gispen *et al*, 1980).
2. From the reports published, no firm conclusion regarding an antipsychotic effect for these endorphins can be reached (Manchanda & Hirsch, 1982). The early papers report remarkable clinical effects and were published in a prestigious psychiatric journal (Verhoeven *et al*, 1979, 1982). Since that time, only a limited number of clinical investigations on these compounds have been conducted (Emrich *et al*, 1980; Tamminga *et al*, 1981; Manchanda & Hirsch, 1981, 1982, 1986; Meltzer *et al*, 1982) and the results are divergent. In our opinion, the diverse findings are partly due to variability in the criteria employed for evaluating improvement.

Published results of 'improvement'

A range of results has been reported from centres using DTγE in acute-on-chronic, and chronic schizophrenic patients with persistent florid symptoms (Verhoeven *et al*, 1979; Emrich *et al*, 1980; Tamminga *et al*, 1981; Manchanda & Hirsch, 1981; Meltzer *et al*, 1982). In the various samples of schizophrenic patients treated with this endorphin, the highest percentage of patients showing overall improvement was 78% in two cohorts in Utrecht investigated by Verhoeven *et al* (1979), while the lowest figure was 0% in a study in Baltimore by Tamminga *et al* (1981). Intermediate figures of 46, 25, and 18% were reported by Emrich *et al* (1980) in Munich, Meltzer *et al* (1982) in Chicago, and Manchanda & Hirsch (1981) in London, respectively. However, the initial impression of a wide variation in response between the centres may be misleading.

In the original study by Verhoeven *et al* (1979), 14 schizophrenic patients were assessed using an arbitrary, 3-point (0–2) rating scale, comprising the following items: hallucinations; delusions; train of thought; emotional flattening; orientation; and motor activity. When reporting their results, the investigators used only the change in scores on two psychotic symptoms, namely, delusions and hallucinations.

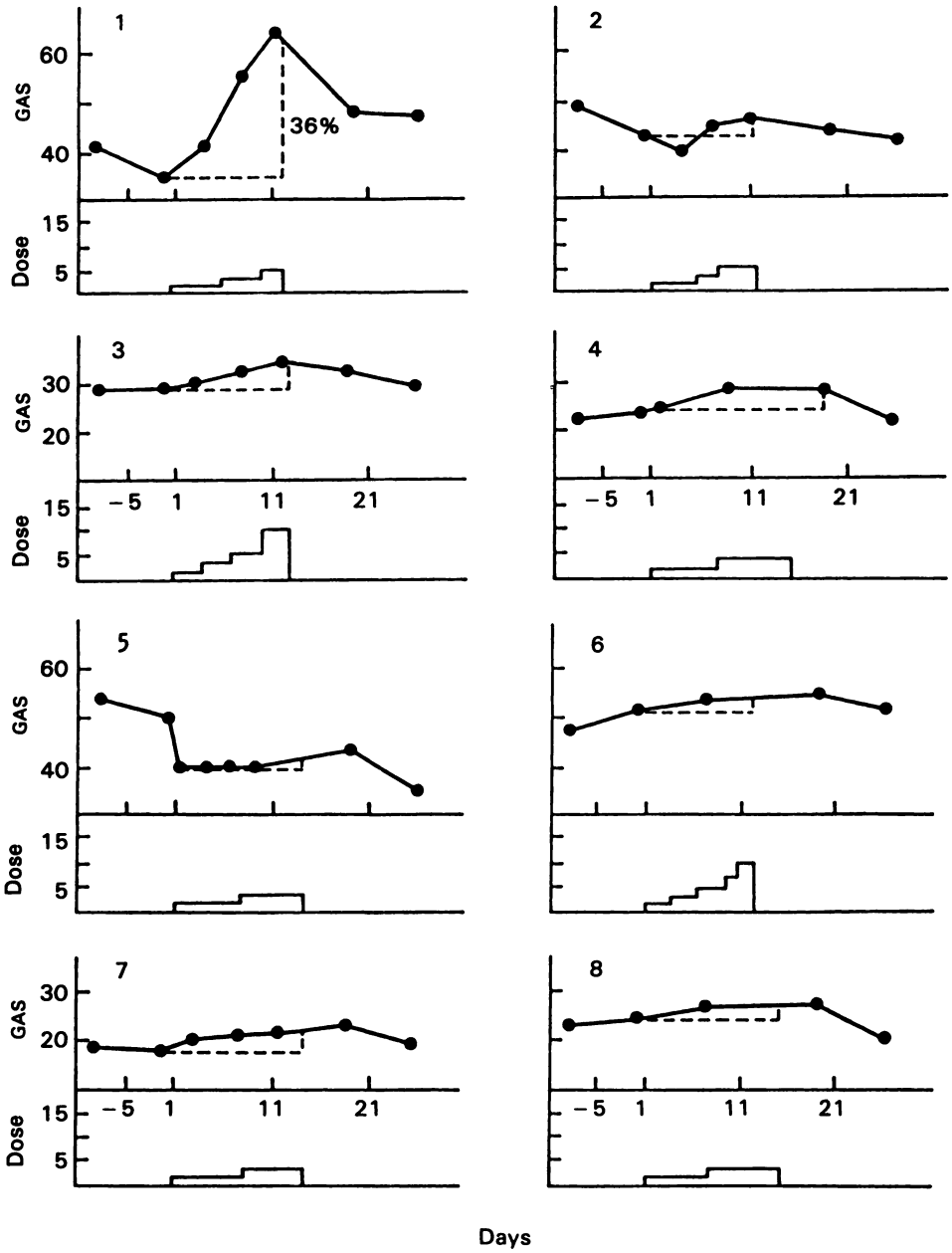

FIG. 1 Global Assessment Scale (GAS) rating in eight patients before, during, and after treatment with DTγ endorphin (from Meltzer *et al*, 1982). Note that a real improvement is seen in patient 1 only. In all other patients, the overall change is only slight. The percentage improvement is indicated by a dotted line (reprinted with the authors' permission).

Further, the authors reported the complete disappearance of these psychotic symptoms in one patient (see Fig. 1, patient 1), in whom mutism and stereotyped behaviour persisted. The rating of the absence of delusions or hallucinations was based solely on the observation of behaviour, and is thus somewhat suspect. Despite these shortcomings in measuring change, claims of a "reduction or total

https://doi.org/10.1192/bjp.153.3.354 Published online by Cambridge University Press

disappearance of symptoms in schizophrenia partly or completely resistant to conventional neuroleptics, and suffering from long-lasting psychosis'' were put forward at the first presentation of the results at the World Congress of the Biological Society in Barcelona in 1978, and in subsequent publications (De Weid *et al*, 1978*b*; Verhoeven *et al*, 1979). This was more than sufficient to arouse international interest in this compound. Subsequent investigations were aimed primarily at replicating the claimed rapid onset of an antipsychotic effect, that was sustained after discontinuation of treatment and free from any serious side-effects.

The results were based on two samples of six and eight patients. Subsequently, another cohort of patient was included, and findings for a total group of 23 schizophrenic patients treated with DTγE were published (Verhoeven *et al*, 1981). For the analysis of these data, the investigators utilised a different criterion for measuring change from that used for the small samples. The response to the treatment with endorphin was assessed by calculating, for each patient, the difference between the total scores on the Brief Psychiatric Rating Scale (BPRS) (Overall & Gorham, 1962) before and after the experimental treatment. This difference was expressed as a percentage of the baseline score, and allowed each patient to be allocated to one of four outcome groups: seven patients showed 'no response' (less than 20% improvement); six patients showed 'slight-to-moderate response' (20–50% improvement); three patients showed 'moderate-to-marked response' (50–80% improvement); and seven patients showed a 'marked response' (more than 80% improvement). Thus, although in the first double-blind study, all eight patients had been reported as showing 'marked improvement' on a 3-point scale, subsequent results from the same investigators on a larger sample, including the original eight patients, revealed that less than one-third of the total sample (7 out of 23 patients) showed marked improvement.

In an uncontrolled study by Meltzer *et al* (1982), eight schizophrenic patients were treated with DTγE for 12 days. The instruments used to evaluate change were the Global Assessment Scale (GAS) (Endicott *et al*, 1976), the SANS-C (Spitzer & Endicott, 1977) with additional items from the Present State Examination (PSE) (Wing *et al*, 1974) for rating psychotic symptoms; and the Nurses Observation Scale for Inpatient Evaluation (NOSIE) (Honigfeld & Klett, 1965); and a scale of global clinical change scored at a weekly ward research meeting. The investigators concluded that only two out of the eight patients showed marked overall improvement. However, the number of patients considered to have improved depends upon the criteria used. For example, on the basis of the SANS-C syndrome score, 57 or 71% could be regarded as showing improvement depending on which symptom is adopted for measuring change, delusions of reference, and persecution or incomprehensibility, respectively (see Table I). The classification of outcome is limited to 'improved', 'no change' or 'deteriorated' categories. Thus, a patient showing even the slightest improvement will qualify for the 'improved' category. Further, using the GAS, they reported that 87% of their patients showed improvement. Close examination of the data for the individual patients, as plotted in Fig. 1, reveals that seven of the eight patients show only very slight improvement. The average mean change of 6%, referred to as a 'trend' by Meltzer *et al* ($P=0.066$) is largely accounted for by a change of about 36% in one patient (Fig. 1, patient 1). When clinicians' assessments were used, two patients showed marked improvement, one showed moderate improvement, three showed minimum improvement, and two showed no change. If any measure of improvement, on any scale, were to be the criterion for response, then 75% of patients would be classified as improved, but this would hardly reflect a clinically useful response to the drug in this study. The overall

TABLE I
Effect of DTγE on SADS-C syndrome score (after Meltzer et al, *1982)*

	Number of subjects with usable ratings	*Improved* n *(%)*	*No change* n *(%)*	*Deteriorated* n *(%)*
Auditory hallucinations	5/8	1 (20)	4 (80)	0 (0)
First-rank symptoms	5/8	2 (40)	3 (60)	0 (0)
Delusions of reference and persecution	7/8	4 (57)	1 (14)	2 (29)
Severity of hallucinations	7/8	2 (29)	4 (57)	1 (14)
Severity of delusions	6/8	0 (0)	2 (33)	4 (67)
Incomprehensibility	7/8	5 (71)	1 (14)	1 (14)

TABLE II
Percentage of patients improved vs percentage change in score

Degree of improvement (Percentage change in rating scale score)	*Utrecht (n = 23)*	*London (n = 11)*	*Munich (n = 13)*	*Chicago (n = 8)*
<20% (no change)	30.5	56	46	25
20–80% (mild–moderate)	39	27	38.5	50
<80% (marked)	30.5	18	15.5	25
Rating scale used	BPRS,3 point	PSE,BPRS,MS	IMPS,VBS	GAS

Table II shows a similarity in outcome when >80% change in scores is utilised as the criteria for improvement.

conclusion of Meltzer *et al* (1982) was that they had failed to replicate Verhoeven's findings.

In our own study (Manchanda & Hirsch, 1981), only two out of eleven schizophrenic patients treated with DTγE showed 'marked improvement' in terms of a statistically significant change in the total PSE scores and the non-specific symptom sub-scores of the Manchester Scale (Krawiecka *et al*, 1977). The psychosis subscores failed to show any appreciable change.

What should be the criteria for improvement?

These examples raise the fundamental question of what should be regarded as a satisfactory criterion for change when evaluating the therapeutic effects of a new drug. In Table II, we compare the results of three studies using standard categories for judging outcome, based on change in rating-scale scores. The method of assessment in our own study (Manchanda & Hirsch, 1981) has been modified in line with the others. Combining patients in both 'slight-to-moderate' and 'marked' improvement categories, it can be concluded that between 36% and 69% of our patients improved with DTγE. In the absence of a control group, further conclusion is limited.

The Utrecht study (Verhoeven *et al*, 1979) included a control group, and reported that the patients on placebo showed a mean improvement of 20%. The Munich group (Emrich *et al*, 1980) also used a placebo-treated control group, and reported that 46% of both groups showed less than 20% improvement. Only 15% of patients in the drug group showed 80% improvement, compared with 31% of the placebo group.

Preliminary, uncontrolled studies with small numbers of patients are often the only testing ground for new drug treatments, and the conclusions of these early investigations can have a powerful influence on the direction, nature, and number of further studies of the drug. One way of helping to ensure valid findings is to employ stringent criteria. Reporting the response of patients in terms of improvement or no improvement may be particularly likely to produce misleading results, unless a positive change is defined in terms of marked improvement, i.e. more than 75–80% change in rating-scale scores. Furthermore, if the response to a drug treatment is to be presented in terms of changes in scores on individual rating-scale items, then these items should be specified, so that it is possible to compare results between centres that have used similar rating scales.

Conclusion

When evaluating a new drug in a clinical trial, differences in the criteria used for measuring change at various centres, and within the same centre, can affect the overall conclusions with respect to outcome. In such studies, we recommend the application of a standard criterion for improvement of at least 80% improvement on a rating-scale score to provide consistent and comparable results that are clinically relevant. Using this criterion, the results from the studies testing DTγE are less discrepant than appears from the original reports (see Table II); in the Utrecht study, 30.5% of patients showed marked improvement, compared with 25% in Chicago, 15.5% in Munich, and 18% in London.

Acknowledgements

The authors thank Professor H. Y. Meltzer and his colleagues and Elsevier Biomedical Press, Amsterdam, Holland for their kind permission to reproduce Fig. 1 and Table I from their article "Effect of des-tyr-gamma endorphin in schizophrenia", published in *Psychiatry Research* (1982), 6, 313–326.

References

DE WEID, D., KOVACS, G. L., BOHUS, B., VAN REE, J. M. & GREVEN, H. M. (1978*a*) Neuroleptic activity of the neuropeptide

-LPH 62-77 (des-tyr-γ-endorphin: DTγE). *European Journal of Pharmacology*, **49**, 427–436.

——, VAN REE, J. M., VERHOEVEN, W. M. A. & VAN PRAAG, H. M. (1978*b*). Endorphins and schizophrenia. Abstract No. 430, *IInd World Congress of Biological Psychiatry*, 31 August–6 September 1978, Barcelona. New York: Plenum.

ENDICOTT, SPITZER, R. L., FLEISS, J. L. & COHEN, J. (1976) The global assessment scale, a procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry*, **33**, 766–771.

EMRICH, H. M., ZANDIG, M., KISSLING, W., DIRLICH, G., ZERRSEN, D. V. & HERZ, A. (1980) Des-tyrosyl-γ-endorphin in schizophrenia, a double blind trial in 13 patients. *Pharmakopsychiatrie*, **13**, 290–298.

GISPEN, W. H., ORMOND, D., TEN HAAF, J. & DE WEID, D. (1980) Modulation of ACTH induced grooming by (Des-tyr)-γ-endorphin and haloperidol. *European Journal of Pharmacology*, **63**, 203–207.

HONIGFELD, G. & KLETT, C. J. (1965) The nurses' observation scale for inpatient evaluation. *Journal of Clinical Psychology*, **21**, 65–71.

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

LOEBER, J., VERHOEF, J., BURBECH, J. P. H. & VAN REE, J. M. (1979) Endorphins and related peptides in human cerebrospinal fluid. *Acta Endocrinology Copenhagen*, **91** (suppl. 225), 74.

MANCHANDA, R. & HIRSCH, S. R. (1981) (Des-tyr^1)-γ-endorphin in the treatment of schizophrenia. *Psychological Medicine*, **11**, 401–404.

—— & —— (1982) (Des-tyr^1)-γ-endorphin for schizophrenia – critical review. *International Pharmacopsychiatry*, **17**, 147–152.

—— & —— (1986) (Des-tyr^1)-γ-endorphin in the treatment of depression: double blind placebo controlled trial. *Human Psychopharmacology*, **1**, 99–102.

MELTZER, H. Y., BUSCH, D. A., TRICOU, B. J. & ROBERTSON, A. (1982) Effect of (des-tyr)-gamma-endorphin in schizophrenia. *Psychiatry Research*, **6**, 313–326.

OVERALL, L. J. E. & GORHAM, D. R. (1962) The brief psychiatric rating scale. *Psychological Reports*, **10**, 799–812.

SPITZER, R. L. & ENDICOTT, J. (1977) *Schedule for Affective Disorders and Schizophrenia (SADS-C)* (3rd edn). New York: New York State Psychiatric Institute.

TAMMINGA, C. A., TIGHE, P. J., CHASE, T. N., DE FRAITES, G. & SCAFFER, M. H. (1981) Des-tyrosine-γ-endorphin administration in chronic schizophrenics – a preliminary report. *Archives of General Psychiatry*, **38**, 167–168.

VERHOEF, J., LOEBER, J. G., BURBACH, J. P. H., GISPEN, W. H., WITTER, A. & DE WEID, D. (1980) α-endorphin, γ-endorphin and their des-tyrosine fragments in rat pituitary and brain tissue. *Life Sciences*, **26**, 851–859.

VERHOEVEN, W. M. A., VAN PRAAG, H. M., VAN REE, J. M. & DE WEID, D. (1979) Improvement of schizophrenic patients by treatment with des-tyr-γ-endorphin. *Archives of General Psychiatry*, **36**, 294–302.

——, VAN REE, J. M., VAN BENTUM, A. H., DE WEID, D., VAN PRAAG, H. M. (1982) Antipsychotic properties of des-enkephalin-γ-endorphin in treatment of schizophrenic patients. *Archives of General Psychiatry*, **39**, 648–654.

——, WESTENBERG, H. G. M., GERRITSEN, A. W., VAN PRAAG, H. M., THIJSSEN, J. H. H., SCHWARZ, F., VAN REE, J. M. & DE WEID, D. (1981) Des-tyrosine-γ-endorphin in schizophrenia. Clinical, biochemical and hormonal aspects. *Psychiatry Research*, **5**, 293–310.

WING, J. K., COOPER, J. E. & SARTORIUS, N. (1974) *The Measurement and Classification of Psychiatric Symptoms*. London: Cambridge University Press.

Rahul Manchanda, MB, BS, MD, FRCP(C), MRCPsych, *Assistant Professor of Psychiatry, University of Western Ontario and Director of Education, St Thomas Psychiatric Hospital, Ontario, Canada*; *Steven R. Hirsch, BA, MD, MPhil, FRCPsych, *Professor of Psychiatry*; Thomas R. E. Barnes, MB, BS, MRCPsych, *Senior Lecturer in Psychiatry, Charing Cross and Westminster Medical School*

*Correspondence: *Charing Cross and Westminster Medical School, Fulham Palace Road, London W6 8RF*