

A COMPARISON OF DOCTOR AND
PATIENT IMPROVEMENT RATINGS IN A
DRUG (MEPROBAMATE) TRIAL

BY

LEE C. PARK, EBERHARD H. UHLENHUTH,
RONALD S. LIPMAN, KARL RICKELS
and
SEYMOUR FISHER

Reprinted from

THE BRITISH JOURNAL OF PSYCHIATRY

Vol. 111, No. 475, June 1965

A Comparison of Doctor and Patient Improvement Ratings in a Drug (Meprobamate) Trial*

By LEE C. PARK, EBERHARD H. UHLENHUTH, RONALD S. LIPMAN,
KARL RICKELS and SEYMOUR FISHER

INTRODUCTION

Out-patient drug trials have relied largely on observer-rated improvement measures to evaluate the efficacy of psychotropic medication (4). Less frequent use (5, 7, 9) has been made of patient ratings or the combination of observer plus patient ratings. Thus a comparison of improvement ratings made by the patient and his treating doctor would be of considerable methodological interest.

The present paper reports such a comparison employing data collected in the course of a double-blind placebo-controlled meprobamate trial (2). The following questions were asked of the data:

- (a) How closely do patients and their doctors agree in rating patients' distress and improvement?
- (b) How closely do different criterion measures agree in rating patients' distress and improvement?
- (c) Are drug-placebo differences equally apparent in both doctor and patient ratings?

SETTING, SUBJECTS AND CRITERION MEASUREMENTS

Three clinics contributed patients: The Henry Phipps Psychiatric Clinic of The Johns Hopkins University, the Psychiatric Clinic of the University of Pennsylvania and the Psychiatric Clinic at Philadelphia General Hospital. Individuals selected were anxious, tense neurotics without sociopathy, organic impairment, alco-

holism or marked depression. The data presented are based on 138 patients, 65 per cent. female, 56 per cent. Negro, mean age 34, who completed a full six-week drug trial with adherence to protocol (2).

Patients had bi-weekly appointments (Weeks 0, II, IV and VI) and filled out criterion forms at each visit immediately before seeing their assigned doctors. The fifteen doctors (psychiatric residents) participating in the study completed the same criterion ratings independently immediately after each patient visit. Patients were asked to rate themselves with respect to their condition during the last week (including day of interview), while doctors were instructed to base their ratings on information obtained in the interview itself, which usually lasted less than 30 minutes.

Three improvement measures are reported:

(a) *Global Improvement*. This is measured by a seven-point scale ranging from seven, "very much worse", through four, "no change", to one, "very much better".

(b) *Target Symptoms*. This criterion measure is derived from a 64-item modified Hopkins Symptom Checklist (Discomfort Scale) covering common psychic and somatic complaints (3, 6). The patient is asked to check how much each symptom has bothered him under the categories "not at all", "a little", "quite a bit", and "extremely", scored 1-4.

In addition to these categories, a "not elicited" category is present on the doctor's form. The doctor is instructed not to infer symptoms but to check "not elicited" for any symptoms not actually mentioned by the patient. Target Symptoms are defined at the first interview as any complaints checked as

* These data were collected as part of an investigation supported by two NIMH-PSC research grants from the National Institute of Mental Health to The Johns Hopkins University (MH-04732) and to the University of Pennsylvania (MH-04731).

present by *both* the doctor and patient. Thus in effect, Target Symptoms represent an attempt to individualize the particular symptoms that are so salient they are reported by the patient both on a paper form and verbally to the doctor.

An average of 11.59 Target Symptoms were defined per patient in the sample. The mean intensity of each patient's Target Symptoms was followed through the course of treatment.

(c) *Anxiety Scale.* The Anxiety Scale is made up of 15 adjectives, 10 weighted positively and 5 negatively, which tap nervousness, tension, worry, etc. As with Target Symptoms, there are four categories ranging from "not at all" to "extremely", and in addition, the physician has an "unrateable" category. The doctor is instructed to infer symptoms if he feels there is sufficient evidence.*

RESULTS

Agreement Between Patients and Doctors in Assessment of Distress and Improvement

In Table I the mean improvement scores are presented by rater for Global Improvement at Weeks II, IV and VI. It is clear from Table I that patients and doctors are in very close agreement regarding the amount of improvement that has occurred since the start of treatment. None of the doctor and patient means are reliably different at any time period. Thus, after two weeks of treatment, patients and doctors agree that the patients are feeling, roughly, "a little better". At Week VI, patients are on the average feeling somewhere between "a little better" and "quite a bit better".

TABLE I

Doctor and Patient Ratings of Mean Global Improvement Post-Medication Week*

Rater	II	IV	VI
Patient	2.92	2.60	2.74
Doctor	2.94	2.78	2.64

* Note: *N* varies from 132 to 138.

* Another improvement measure, a Depression Scale, was also administered, but the scores have not been evaluated. Primarily depressed patients were excluded from the study.

Target Symptom and Anxiety Scale ratings are shown in Table II. Analysis of these data reveals rather different findings from Global Improvement, which cannot be rated at Week 0. Doctor and patient Target Symptom ratings show a reliable mean difference in premedication distress level (Week 0, $t=3.54$, $p<.001$). At subsequent time periods mean differences are not reliable. Tests of mean Target Symptom Improvement between doctor and patient ratings from Week 0 to Weeks II, IV and VI are significant at Week II ($t=2.06$, $p<.05$) and approach significance for Week VI ($t=1.66$, $.10>p>.05$).

Doctor and patient Anxiety Scale comparisons also show a reliable difference in mean premedication distress level (Week 0, $t=3.58$, $p<.001$). At subsequent time periods mean distress level differences are not reliable. Tests of mean Anxiety Scale Improvement between doctor and patient ratings are significant from premedication to Weeks II, IV and VI (Week II, $t=2.12$, $p<.05$; Week IV, $t=3.31$, $p<.001$; Week VI, $t=2.83$, $p<.01$).

TABLE II

*Doctor and Patient Ratings of Mean Target Symptom Distress Level**

Rater	Pre-Medication Level (Week 0)	Post-Medication Week		
		II	IV	VI
Patient ..	2.70	2.28	2.16	2.13
(change)		(0.42)	(0.54)	(0.57)
Doctor ..	2.90	2.35	2.27	2.20
(change)		(0.55)	(0.63)	(0.70)

*Doctor and Patient Ratings of Mean Anxiety Distress Level**

Rater	Pre-Medication Level (Week 0)	Post-Medication Week		
		II	IV	VI
Patient ..	.32	.02	-.04	-.19
(change)		(.30)	(.36)	(.51)
Doctor ..	.77	.22	.02	-.10
(change)		(.55)	(.75)	(.87)

* Note: *N* varies from 132 to 138.

Table III shows Pearson's r 's between doctor and patient distress ratings (Target Symptoms and Anxiety at each visit) and between doctor and patient Anxiety, Target Symptom and Global Improvement ratings (pre-post differences), respectively, for post-medication Weeks II, IV and VI.

TABLE III
Correlation of Doctor and Patient Distress Level and Improvement Ratings*

A.—Distress Level			
	Dr. Pt. Anx.	Dr. Pt. T/S	
Week 0	.27	.06	
Week II	.33	.54	
Week IV	.32	.50	
Week VI	.49	.56	

B.—Improvement (Pre-Post Scores)			
	Dr. Pt. Anx.	Dr. Pt. T/S	Dr. Pt. Global
Week II	.21	.55	.48
Week IV	.25	.47	.46
Week VI	.38	.51	.43

* Note: All correlations are significant but $\cdot 06$.

The magnitude of the correlations between patient and doctor distress ratings is lower at Week 0 than at subsequent weeks and is not reliably different than zero ($r = \cdot 06$) for Target Symptoms at Week 0. Generally, the magnitude of these correlations increases as a function of increased doctor-patient contact.

Improvement correlations are all positive and significantly different than zero at better than the $\cdot 05$ level of confidence. Thus, it can be seen that there is reliable agreement regarding the rank ordering of patient improvement on the same criterion measures by patients and their doctors.

Finally, it should be noted that there is a consistent trend for doctor and patient Anxiety Scale improvement ratings to be in less agreement than doctor and patient Global and Target Symptom Improvement ratings. Tests of the difference between r 's (Edwards, p. 131-132) show these differences to be significant

at Week II ($p < \cdot 05$) and Week IV ($p < \cdot 05$) but not at Week VI (1). This same trend can be seen in comparing doctor and patient Anxiety and Target Symptom distress level ratings.

Agreement Between Different Criterion Measures of Distress and Improvement

As shown in Table IV, there is significant agreement between different measures with regard to rank ordering of distress and improvement. This is true for criterion measures of distress and improvement (a) as observed by doctors and (b) as experienced and reported by patients.

Patient criterion ratings are somewhat more internally consistent than are doctor ratings, particularly with regard to Anxiety. Both patient and doctor criterion ratings are generally more highly correlated within rater type than they are across rater type. The Mann Whitney U-Test (8) was employed to compare the magnitude of the 9 doctor-patient improvement correlations across the same criterion-measure (Table III, Section B) with the 9 doctor-doctor improvement correlations across different criterion measures (Table IV, Section A) and also with the 9 patient-patient improvement correlations (Table IV, Section B). These data have been cast in a contingency format in Table V for reader examination.

Patient ratings are more highly correlated ($p < \cdot 02$) across the different criterion measures than they are with doctor ratings on the same criterion measures. This same trend (p roughly $\cdot 10$) was found for doctor-doctor *vs.* doctor-patient ratings.

Drug Effects: Improvement Ratings of Doctors and Patients

There was some slight tendency for main drug effects to be more apparent in patient than in doctor ratings.*

* In addition to medication, a second independent patient expectation or "set" variable was also included in this study (2). Its influence, however, is not germane to the contents of this report.

TABLE IV

*Correlation of Criterion Measures of Distress and Improvement**

A.—Doctor Ratings										
					Anx. vs. T/S		Anx. vs. Global		T/S vs. Global	
					Distress (Imp.)		Distress (Imp.)		Distress (Imp.)	
Week 036	—	—	—	—	—
Week II46	(.43)	.44	(.46)	.57	(.59)
Week IV39	(.44)	.46	(.54)	.51	(.51)
Week VI52	(.54)	.49	(.49)	.56	(.51)

B.—Patient Ratings										
					Anx. vs. T/S		Anx. vs. Global		T/S vs. Global	
					Distress (Imp.)		Distress (Imp.)		Distress (Imp.)	
Week 045	—	—	—	—	—
Week II71	(.71)	.39	(.42)	.53	(.51)
Week IV74	(.76)	.53	(.49)	.57	(.60)
Week VI74	(.73)	.48	(.51)	.50	(.51)

* Note: All correlations are significant beyond the .05 level of confidence.

In general, doctor criterion ratings tended to reveal drug effects early and patient ratings late in the trial. The Global Improvement and Target Symptom criterion scales proved more sensitive to drug-placebo differences than did the Anxiety Scale.

TABLE V

Improvement Rating Correlations (from Table III and Table IV) Cast in a Contingency Table Format

Patient with Patient vs. Patient with Doctor			
	Above Median	Below Median	Total
Patient/Patient ..	7	2	9
Patient/Doctor ..	2	7	9
	9	9	18

Doctor with Doctor vs. Doctor with Patient			
	Above Median	Below Median	Total
Doctor/Doctor ..	6	3	9
Doctor/Patient ..	3	6	9
	9	9	18

DISCUSSION

The correlations between improvement ratings by patients and by their doctors and the correlations between improvement ratings on different criterion measures are generally modest, though significant. They suggest (a) substantial differences in the aspects of improvement measured by the various raters and measures, and possibly (b) large errors in the ratings.

The fact that the correlations among the different measures for the same rater are higher than the correlation between raters for the same measure suggests there are systematic differences in the aspects of improvement measured by the patient and the doctor. Thus, although there is modest rater agreement there also seems to be some *real* (not attributable to error variance) uniqueness in the manner in which each rater type perceives patient improvement.

The means data on Target Symptoms and Anxiety (Table II) suggest two characteristic differences between ratings made by patients

and their doctors: (1) the doctors' ratings reveal greater patient improvement (a larger decrease in distress level from pre- to post-treatment) and (2) doctors initially rate their patients as more distressed than do the patients themselves. In keeping with this latter point, it will be recalled that distress level correlations (Table III) were also lower at the first appointment than at subsequent visits. Thus, it seems clear that the doctor is generally less in tune with the patient initially than after he has had more extended contact with the patient. The higher doctor initial distress ratings may relate to the fact that the initial appointment tends to involve a distress-focused interview; since subsequent appointments focus on improvement and assets as well as on distress, the doctor may then develop a more accurate impression of the patient. In addition, in an unpublished study by the senior author using these same measures but more experienced doctors, this initial doctor-patient rating difference was much less marked.

Doctors' and patients' ratings do not show clear differences in sensitivity to drug effects, although we suspect patient ratings may have the edge.

The generally lower concordance between raters on the Anxiety Scale may reflect the fact that the doctor was instructed to base Global and Symptom Checklist ratings on the verbal report of the patient, whereas his anxiety rating was supposed to be a more inferential judgment.

In conclusion, we would speculate that our results, with regard to different raters and criterion measures, are reflecting the complexity and multidimensionality of out-patient response to pharmacotherapy. It would seem that both patient and doctor ratings, each sensitive to somewhat different aspects of the treatment situation, may give the most comprehensive clinical picture.

SUMMARY

This is a comparison of doctor and patient ratings on three improvement measures in a six-week drug study involving 138 anxious, neurotic

out-patients seen at The Johns Hopkins Hospital, The Hospital of the University of Pennsylvania and the Philadelphia General Hospital.

At the premedication interview, doctors rate their patients as significantly more distressed than the patients rate themselves on a Target Symptom measure and an Anxiety Scale. The correlations between doctor and patient ratings are particularly low at this interview.

At subsequent interviews, doctors and their patients do not differ reliably in rating the patients' distress on the Target Symptom and Anxiety measures. The raters also do not differ reliably on a Global Improvement measure (not scored at premedication interview).

The degree of agreement among the three measures in rating improvement (Global Improvement, pre-post difference in Target Symptoms and Anxiety) also was examined by means of correlations. These correlations are all significant, although generally of modest magnitude, at the 2, 4 and 6 week interviews. The correlations among the three ratings made by the patients are about the same as the correlations among the three ratings made by the doctors. (The correlations between raters on the same measures, however, are somewhat lower, particularly for the Anxiety Scale.)

Doctors' and patients' ratings are about equally sensitive in detecting drug-placebo differences. The Anxiety Scale appears less sensitive to these differences than the other two measures.

The use of several measures and different raters seems to offer advantages in tapping various facets of change in patients' distress.

REFERENCES

1. EDWARDS, ALLEN L. (1950). *Experimental Design in Psychological Research*. New York: Rinehart and Co., Inc.
2. FISHER, S., COLE, J. O., RICKELS, K., and UHLENHUTH, E. H. (1962). "Drug-set inter-action: the effect of expectations on drug response in out-patients." In *Neuro-Psychopharmacology* (eds. Bradley, P. B., Flügel, F., and Hoch, P.) Amsterdam: Elsevier Pub. Co.
3. FRANK, JEROME D., GLIEDMAN, LESTER H., IMBER, STANLEY D., and NASH, EARL H. (1957). "Why patients leave psychotherapy." *A.M.A. Arch. Neurol. Psychiat.*, 77, 283-299.

4. LORR, MAURICE (1960). "Rating scales, behavior inventories and drugs", in *Drugs and Behavior* (eds. Uhr, Leonard and Miller, James G.). New York and London: John Wiley and Sons, Inc.
5. —, McNAIR, DOUGLAS M., and WEINSTEIN, GEORGE J. (1963). "Early effects of chlordiazepoxide (Librium) used with psychotherapy." *J. Psychiat. Res.*, **1**, 257-270.
6. PARLOFF, MORRIS B., KELMAN, HERBERT C., and FRANK, JEROME D. (1954). "Comfort, effectiveness and self-awareness as criteria of improvement in psychotherapy." *Amer. J. Psychiat.*, **111**, 343-351.
7. RICKELS, KARL, CLARK, THOMAS W., EWING, JAMES H., and KLINGENSMITH, WALTER C. (1959). "Evaluation of tranquilizing drugs in medical out-patients." *J.A.M.A.*, **171**, 1649-1656.
8. SIEGEL, SIDNEY (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill Book Company, Inc.
9. UHLENHUTH, E. H., and PARK, LEE C. (1964). "The influence of medication (imipramine) and doctor in relieving depressed psycho-neurotic out-patients." *J. Psychiat. Res.*, **2**, 101-122.

Lee C. Park, M.D., *Assistant Professor of Psychiatry*

Eberhard, H. Uhlenhuth, M.D., *Assistant Professor of Psychiatry*
Johns Hopkins University School of Medicine, Baltimore, Maryland

Ronald S. Lipman, Ph.D., *Research Psychologist, Psychopharmacology Service Center, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland*

Karl Rickels, M.D., *Associate Professor of Psychiatry, School of Medicine, University of Pennsylvania, and Director of Psychopharmacological Research, Philadelphia General Hospital, Philadelphia, Pennsylvania*

Seymour Fisher, Ph.D., *Research Professor and Director, Psychopharmacology Laboratory, Division of Psychiatry, Boston University School of Medicine, Boston, Massachusetts*

(Received 4 August, 1964)



HEADLEY BROTHERS LTD
109 Kingsway London WC2
and Ashford Kent