Early Improvement Under Mirtazapine and Paroxetine Predicts Later Stable Response and Remission With High Sensitivity in Patients With Major Depression

Armin Szegedi, M.D.; Matthias J. Müller, M.D.; Ion Anghelescu, M.D.; Christoph Klawe, M.D.; Ralf Kohnen, Ph.D.; and Otto Benkert, M.D.

Received Jan. 2, 2002; accepted Aug. 8, 2002. From the Department of Psychiatry, University of Mainz, Mainz (Drs. Szegedi, Müller, Anghelescu, Klawe, and Benkert), and the Institute for Medical Research Management and Biometrics (IMEREM), Nürnberg, Germany (Dr. Kohnen).

The clinical trial from which these results have been derived was supported by a grant from Organon GmbH, Munich, Germany. The analyses presented here were carried out independently.

The results have been presented as an oral communication at the World Congress of Biological Psychiatry, Berlin, 2001, and at the 22nd meeting of the Association of Neuropsychopharmacology and Pharmacopsychiatry (AGNP), Nürnberg, 2001.

Corresponding author and reprints: Armin Szegedi, M.D., Department of Psychiatry, University of Mainz, Untere Zahlbacher Str. 8, D- 55131 Mainz (e-mail: szegedi@mail.psychiatrie.klinik.uni-mainz.de).

Objective: Current clinical knowledge holds that antidepressants have a delayed onset of efficacy. However, the delayed onset hypothesis has been questioned recently by survival analytical approaches. We aimed to test whether early improvement under antidepressant treatment is a clinically useful predictor of later stable response and remission.

Method: We analyzed data from a randomized double-blind controlled trial with mirtazapine and paroxetine in patients with major depression (DSM-IV). Improvement was defined as a 17-item Hamilton Rating Scale for Depression (HAM-D-17) score reduction of ≥ 20%. Stable response was defined as ≥ 50% HAM-D-17 score reduction at week 4 and week 6, and stable remission as a HAM-D-17 score of ≤ 7 at week 4 and week 6. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated.

Results: Improvement occurred in a majority of the analyzed patients within 2 weeks (mirtazapine: 72.7% of 109 patients; paroxetine: 64.9% of 103 patients). Early improvement was a highly sensitive predictor of later stable response or stable remission for both drugs. NPV approached maximum values as early as week 2 for mirtazapine and week 3 for paroxetine. After 2 weeks of treatment with mirtazapine and 3 weeks with paroxetine, almost none of the patients who had not yet improved became a stable responder or stable remitter in the later course.

Conclusion: Our results strongly suggest that early improvement predicts later stable response with high sensitivity. These empirically derived data question the delayed onset hypothesis for both antidepressants tested and provide important clinical clues for an individually tailored antidepressant treatment.

(J Clin Psychiatry 2003;64:413–420)

After several decades of research in antidepressant treatment, reliable predictors of response to a given antidepressant have not yet been identified. Clinicians still face the problem that individualized antidepressant drug selection is guided mainly by trial and error with regard to the desired antidepressant effect. The problem is made worse by the common clinical view that antidepressant response usually appears with a delay of several weeks. This notion stems mainly from 2 sources: first, controlled clinical trials aiming to provide evidence for an antidepressant's efficacy usually compare the active compound with placebo. By comparing mean scores of rating scales as measures for depressive symptomatology using repeated measurement ANOVA, a significant difference between active treatment and placebo usually is detected from week 3 onward.[1,2] Second, pattern analyses conducted by Quitkin and coworkers[3–5] have suggested that persistent or "true" drug response occurs mainly in the later course of treatment, i.e., week 3–4 of treatment, while response occurring in the first 2 weeks was assumed to be unstable and due to placebo effects. Taken together, these findings have had substantial impact on clinical practice and led to the hypothesis of delayed action of antidepressants. For the depressed patient, an adequate treatment trial is thought to last at least 4–6 weeks, until nonresponse can be assumed, which requires substantial patience and compliance, and may imply the risk of fatal complications like suicide attempts.

However, recent publications have questioned the delayed onset hypothesis for antidepressants.[6–9] By analyz-

ing intraindividual courses during treatment by means of survival analytical techniques, Stassen and colleagues[8,10,11] found that patients who finally responded to an antidepressant showed substantial improvement in the early course of treatment (i.e., within the first 2 weeks). Improvement (defined as ≥ 20% score reduction on the Hamilton Rating Scale for Depression [HAM-D]) was predictive of later response (defined as ≥ 50% 17-item HAM-D [HAM-D-17] score reduction).

At the first glance, the results of survival analytical approaches and pattern analyses seem contradictory and lead to different clinical implications. While the results of Quitkin et al.[3–5] lead to the recommendation to prolong treatment up to 6 weeks in order to identify true drug responders, the results of Stassen et al.[8,10,11] postulate that improvement, which occurs early in the course of treatment, predicts response, and if improvement does not occur, there will be little chance of response if the treatment strategy is not changed by the clinician.

This report presents data from a randomized controlled trial comparing mirtazapine and paroxetine in patients suffering from a major depressive episode according to DSM-IV criteria. The clinical data of the trial have been analyzed in order to test the following issues:

- Does improvement occur early in the course of treatment in a majority of patients?
- Does early improvement (within the first 2–3 weeks) predict later stable response or stable remission?
- How sensitive and specific is early improvement as a predictor of later stable response or stable remission? What are the positive predictive value (PPV) and negative predictive value (NPV) of early improvement?
- If improvement does not occur in a patient, when should the clinician change the treatment strategy?

METHOD

A multicenter, randomized, double-blind comparison of mirtazapine and paroxetine was conducted in 50 centers in Germany. The study protocol was approved by the local ethics committees and the study was conducted in accordance with good clinical practice standards and in accordance with the Declaration of Helsinki. All patients provided written informed consent after the procedures and possible side effects had been fully explained.

Patients

Patients were recruited in general practices and in psychiatric outpatient departments. In total, 11 research assistants (psychiatric residents, research fellows, or psychologists) who had been trained in rating patients with the psychiatric scales and in performing structured diagnostic interviews with the Mini International Neuropsychiatric Interview (MINI)[12] were responsible for eligibility check of selected patients. Training was based on the rating of 5 different videotapes from patients with major depressive disorder requiring that the sum scores in the HAM-D-17 of any rater did not differ more than ± 2 points from an independent expert rating; the MINI was taught in a 1-day lecture and trained in 3–5 supervised exercise administrations in volunteers. The research assistants were engaged by a clinical research organization (IMEREM, Nürnberg, Germany) and met the patients at the centers to perform all psychiatric investigations and ratings during the study.

Patients (male or female, aged 18–70 years) fulfilling DSM-IV[13] criteria for major depressive episode and with a total score of ≥ 18 on the HAM-D-17[14] at the start and end of the placebo washout (3–7 days) period were eligible for inclusion in the study. Reasons for exclusion included a current depressive episode of more than 12 months' duration, a lack of response to at least 2 adequate antidepressant therapies during the current episode, more than 3 previous episodes that did not respond to adequate antidepressant therapy, a reduction of ≥ 25% in the HAM-D-17 score during the placebo washout period, suicide risk defined as a Montgomery-Asberg Depression Rating Scale (MADRS)[15] score in item 10 (suicide ideations) of 4 to 6, and current bipolar disorder, depressive disorder not otherwise defined, panic disorder (with or without agoraphobia), agoraphobia without a history of panic disorder, schizophrenia, organic mental disorder, eating disorder (anorexia or bulimia nervosa), specific phobia, social phobia, or generalized anxiety disorder. The latter 3 conditions were only considered as exclusion criteria if they caused clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Patients were excluded if they were suffering from alcohol/substance abuse or epilepsy, if they had a history of seizure disorder, if they had ever received treatment with an anticonvulsant for epilepsy or seizures, or if they had clinically meaningful physical disease or abnormal findings on physical examination or laboratory testing. Female patients were also excluded if they were pregnant, lactating, or of childbearing potential and not taking adequate contraceptive measures. The following treatments must have been stopped within the indicated intervals before the start of active study medication: electroconvulsive therapy (3 months), depot neuroleptics (2 months), fluoxetine (4 weeks), benzodiazepines (2 weeks), monoamine oxidase inhibitors (2 weeks), paroxetine (current episode), and other psychotropic drugs (1 week).

Treatment Schedule

Following a 3- to 7-day placebo washout period, patients were randomly assigned to receive treatment with either mirtazapine or paroxetine for 6 weeks. The dose of

mirtazapine was increased from 15 mg/day on days 1 and 2, to 30 mg/day from day 3 onward; an increase to 45 mg/day or maintenance of the 30-mg/day dose was performed after 2 weeks in nonresponders by randomization. Nonresponders were defined by Clinical Global Impressions (CGI)[16] ratings in item 3 (therapeutic efficacy) of "slight" or "unchanged/worsened" and no "outweighs therapeutic efficacy" rating in item 4 (tolerability). Paroxetine was started at a dose of 20 mg/day. After 2 weeks, nonresponders were randomly assigned either to continue on 20 mg/day or to increase the dose to 40 mg/day until the end of the study. Both drugs were given once daily, mirtazapine in the evening and paroxetine in the morning, using a double-dummy technique.

Concomitant treatment with psychotropic (including benzodiazepines) or sedative (including sedative antihistamines) drugs or the antihypertensive medications guanethidine, guanoxan, clonidine, prazosin, or alpha-methyldopa was not allowed. The only exception was chloral hydrate (1 g/day for up to 3 successive days/week), which was permitted for sleeping problems.

Assessments

Assessments were performed at screening, baseline (day 0), and at weeks 1, 2, 3, 4, and 6 of active treatment or on premature withdrawal. Efficacy was assessed using the HAM-D-17 rating scale.

Statistical Analyses

Improvement was defined as a HAM-D-17 score reduction of ≥ 20% compared with baseline. *Sustained improvement* was defined as a HAM-D-17 score reduction of ≥ 20% compared with baseline that was subsequently maintained at all time points of assessment. *Stable response* was defined as ≥ 50% HAM-D-17 score reduction at week 4 and week 6. *Stable remission* was defined as displaying a HAM-D-17 total score of 7 or less at week 4 and 6. Sensitivity, specificity, PPV, NPV, and receiver operating characteristic (ROC) curves were calculated.[17]

Quantitative data were analyzed using the Student t test. Qualitative data were analyzed using the chi-square test or the Fisher exact test for binary data, or the Mantel-Haenszel chi-square test for ordinal data. All tests were 2-sided and statistical significance was defined as $p \leq .05$. All statistical analyses were performed using SPSS Statistical Software package (SPSS Inc., Chicago, Ill.).

RESULTS

Overall Results

A total of 311 patients were screened for participation in the study, 275 of whom were randomly assigned to treatment (139 in the mirtazapine group and 136 in the paroxetine group). From 6 patients, no postbaseline values for safety or efficacy were available due to early withdrawal from the study (lost to follow-up). In total, 250 were included in the intention-to-treat (ITT) sample (mirtazapine: N = 127; paroxetine: N = 123). From 109 patients of the mirtazapine-treated group and 103 of the paroxetine-treated group, complete data sets were available for the calculations of sensitivities, specificities, PPVs, and NPVs. The number of and reasons for dropouts were similar in both groups.

Table 1. Patients' Characteristics

Characteristic	Mirtazapine (N = 135)	Paroxetine (N = 134)
Age, mean ± SD, y	47.2 ± 11.1	47.3 ± 10.3
Male/female, %	37/63	35/65
First episode of major depression, %	41.5	43.3
Previous episodes of major depression, %	56.3	54.5
Duration of current episode, mean ± SD, d	97.7 ± 86	110.4 ± 104
HAM-D-17, mean ± SD baseline score	22.4 ± 3.3	22.4 ± 3.2

Abbreviation: HAM-D-17 = 17-item Hamilton Rating Scale for Depression.

Both treatment groups were well matched at baseline with respect to demographic and disease characteristics (Table 1).

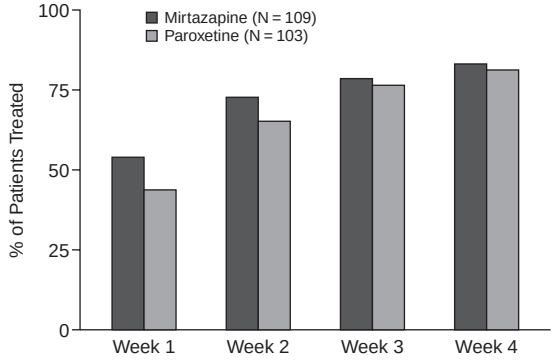
The majority of patients in both groups (55% in the mirtazapine group and 59% in the paroxetine group) were rated as markedly ill on the CGI severity of illness scale; 13% in each group were rated as severely ill.

The mean daily dosage was 32.7 mg of mirtazapine and 22.9 mg of paroxetine. The majority of patients (98 [77.2%] in the mirtazapine group and 94 [76.4%] in the paroxetine group) did not require dose escalation after 2 weeks. Dose escalation was necessary in 23 patients (18.1%) in the mirtazapine group (to 45 mg/day) and 18 patients (14.6%) in the paroxetine group (to 40 mg/day). The remaining 17 patients received the starting doses and withdrew from the study before the end of week 2.

The results with regard to efficacy, reasons for dropouts, and tolerability have been published in detail elsewhere.[18] In summary, both treatments were equally effective in reducing the mean HAM-D-17 total score. However, mirtazapine was associated with a faster onset of action, the mean HAM-D-17 score being significantly lower in the mirtazapine group than in the paroxetine group at week 1 (16.5 vs. 18.8, $p = .0032$). At endpoint, the mean HAM-D-17 total score was 10.7 in the mirtazapine group and 11.9 in the paroxetine group.

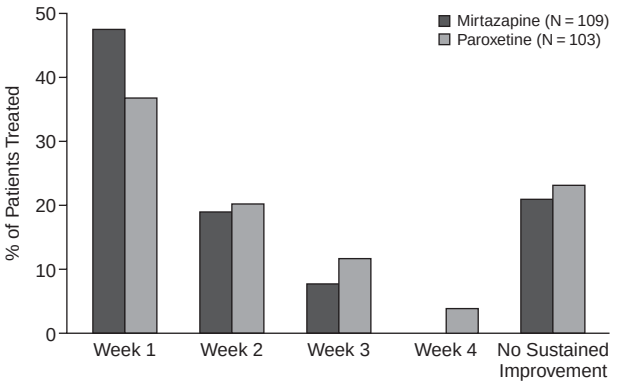
From the mirtazapine group, 69 (63.3%) of 109 patients were considered responders (at least 50% score reduction of HAM-D-17) at week 6; the respective figures were 62 (60.2%) of 103 patients in the paroxetine group. Remission (defined as ≤ 7 points on HAM-D-17) at week 6 was found in 49 (45.0%) of 109 patients for mirtazapine, and in 40 (38.8%) of 103 patients for paroxetine.

Figure 1. Percentage of Improvers (≥ 20% decrease of HAM-D-17 total score) During the First Weeks of Treatment



Abbreviation: HAM-D-17 = 17-item Hamilton Rating Scale for Depression.

Figure 2. Percentage of Sustained Improvers (≥ 20% HAM-D-17 score reduction and consistently maintained), With Sustained Improvement Starting at Week X



Abbreviation: HAM-D-17 = 17-item Hamilton Rating Scale for Depression.

Early Improvement

When the sample was analyzed according to the criterion of improvement (as defined by a decrease of at least 20% of the initial HAM-D-17 total score), we found that during the first 2 weeks of treatment a high percentage of patients in both treatment groups fulfilled this criterion (Figure 1), with numerical advantages for mirtazapine (at week 1: mirtazapine, 54.0%; paroxetine, 43.9%; at week 2: mirtazapine, 72.7%; paroxetine, 64.9%). Only a relatively small proportion of patients additionally fulfilling the improvement criterion were found at weeks 3 and 4 (total percentages week 3: mirtazapine, 78.6%; paroxetine, 76.6%; week 4: mirtazapine, 83.5%; paroxetine, 81.8%; all percentages refer to the sample with complete data sets; mirtazapine, N = 109; paroxetine, N = 103).

In order to ensure that the observed improver rates were not simply due to unstable fluctuations, we additionally analyzed our sample according to a stricter improvement criterion, which demanded that once observed, improvement had to be persistingly present in the later course of treatment (sustained improvement). The results of this analysis, showing the percentage of patients who fulfilled this strict criterion week by week, are plotted in Figure 2. The vast majority of sustained improvements occurred during the first 2 weeks of treatment.

Early Improvement and Prediction of Stable Response

After demonstrating that improvement occurred in the early course of treatment in a majority of patients, we analyzed whether early improvement was a reliable predictor of later stable response. We chose a relatively strict response criterion which required that a patient had to maintain a decrease in the HAM-D-17 total score of at least 50% from baseline at least from week 4 to week 6. This criterion seems reasonable from a clinical point of view.

We calculated sensitivity, specificity, PPV, and NPV of the test (early improvement) to predict later stable response. PPV in this context means: if the predictor (early improvement) is positive, what is the chance that the criterion (stable response) is fulfilled? NPV means in this context: if the predictor (early improvement) is negative, what is the chance that the criterion (stable response) is not fulfilled? The respective figures are shown in Table 2A.

As our results indicate that early improvement is a highly sensitive predictor of later stable response, we tried an empirically derived answer to the clinically important question: how long should the clinician wait until the treatment strategy is changed, if the patient does not improve? In statistical terms, the interesting figures in this context are the percentages of false negatives (i.e., reciprocal values of the sensitivity), which express the percentage of patients not improving by at least 20% at week X, but who later still become stable responders. These figures would give the clinician an estimate about the risk, i.e., how many potential responders the clinician would miss if the clinician would base the clinical decision on the predictor (early improvement). The respective figures are also given in Table 2A. As it can be seen from these data, none of the mirtazapine-treated patients who had *not* improved after 2 weeks and none of the paroxetine patients who had *not* improved after 3 weeks became a stable responder in the later course of treatment.

The predictor of at least 20% improvement was chosen deliberately, because it has been reported to be a useful threshold value, which can be assessed reliably and represents a clinically meaningful change in the patient's state. In addition, we performed the same analyses for other cut-offs of improvement (at least 25% or at least 30% improvement, respectively), because other authors in the literature have used these cut-offs. The results are listed in Table 2B and 2C. As one can see, the data resemble those for the 20% cut-off in general, but the sensitivity tends to

Table 2. Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) of Improvement in HAM-D-17 Score to Predict Later Stable Response (A–C)[a] and Response at Week 6 (D)

A. Improvement (predictor) "at least 20% reduction"

Week	Sensitivity MIR/PAR	Specificity MIR/PAR	PPV MIR/PAR	NPV MIR/PAR	False Positives MIR/PAR	False Negatives MIR/PAR
1	0.79/0.65	0.73/0.67	0.75/0.56	0.77/0.75	0.27/0.33	0.21/0.35
2	0.97/0.91	0.53/0.50	0.69/0.54	0.94/0.89	0.47/0.50	0.03/0.09
3	1.00/0.98	0.42/0.36	0.64/0.49	1.00/0.96	0.58/0.64	0/0.02
4	1.00/1.00	0.35/0.30	0.62/0.48	1.00/1.00	0.65/0.70	0/0

B. Improvement (predictor) "at least 25% reduction"

Week	Sensitivity MIR/PAR	Specificity MIR/PAR	PPV MIR/PAR	NPV MIR/PAR	False Positives MIR/PAR	False Negatives MIR/PAR
1	0.69/0.58	0.80/0.76	0.78/0.61	0.71/0.74	0.20/0.24	0.31/0.42
2	0.88/0.91	0.62/0.56	0.71/0.57	0.83/0.90	0.38/0.44	0.12/0.09
3	1.00/0.95	0.47/0.41	0.67/0.51	1.00/0.93	0.53/0.59	0/0.05
4	1.00/1.00	0.45/0.44	0.66/0.54	1.00/1.00	0.55/0.56	0/0

C. Improvement (predictor) "at least 30% reduction"

Week	Sensitivity MIR/PAR	Specificity MIR/PAR	PPV MIR/PAR	NPV MIR/PAR	False Positives MIR/PAR	False Negatives MIR/PAR
1	0.66/0.40	0.87/0.86	0.84/0.65	0.71/0.69	0.13/0.14	0.34/0.60
2	0.76/0.83	0.66/0.67	0.70/0.61	0.72/0.86	0.34/0.33	0.24/0.17
3	0.98/0.93	0.60/0.52	0.72/0.55	0.97/0.92	0.40/0.48	0.02/0.07
4	1.00/1.00	0.56/0.50	0.71/0.57	1.00/1.00	0.44/0.50	0/0

D. Improvement (predictor) "at least 20% reduction"; response "at least 50% reduction at week 6 only"

Week	Sensitivity MIR/PAR	Specificity MIR/PAR	PPV MIR/PAR	NPV MIR/PAR	False Positives MIR/PAR	False Negatives MIR/PAR
1	0.77/0.57	0.80/0.69	0.87/0.73	0.67/0.52	0.20/0.31	0.23/0.43
2	0.90/0.80	0.45/0.53	0.74/0.71	0.72/0.65	0.55/0.47	0.10/0.20
3	0.99/0.92	0.43/0.41	0.74/0.69	0.94/0.77	0.57/0.59	0.01/0.08
4	0.99/0.95	0.38/0.33	0.73/0.68	0.94/0.82	0.62/0.67	0.01/0.05

[a]Stable response = at least 50% reduction at weeks 4 and 6.
Abbreviations: HAM-D-17 = 17-item Hamilton Rating Scale for Depression, MIR = mirtazapine, PAR = paroxetine.

decrease, the higher the cut-off level chosen. Because the aim of our approach was to identify a predictor with high sensitivity as early as possible in the course of treatment, the cut-off of at least 20% improvement of HAM-D-17 total score seems most appropriate. The results were not substantially influenced by the response criterion. Nearly identical results were obtained when analyzing response at week 6 instead of response at weeks 4 and 6 (stable response) (see Table 2D). When analyzing whether early improvement was mainly due to a subfactor of the HAM-D-17 (Bech melancholia factor, sleep factor, item: "depressed mood"), we found no substantial differences compared to the HAM-D-17 total score analyses (data not shown).

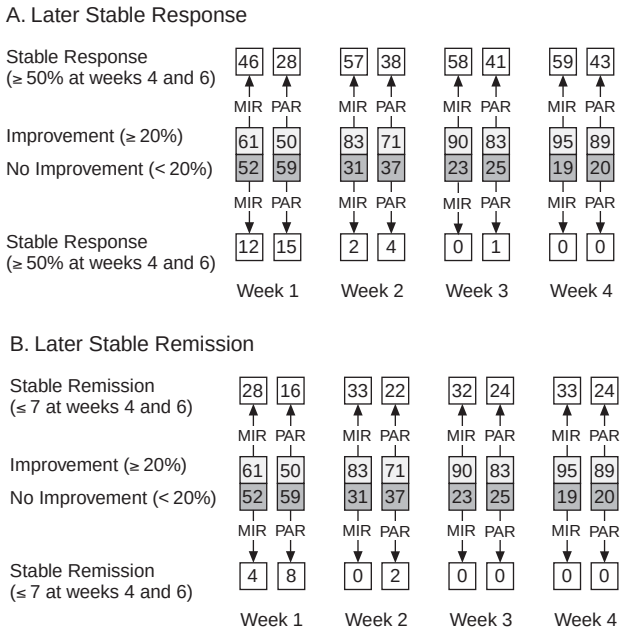
The number of patients with available data sets from the ITT sample who achieved stable response with or without improvement of at least 20% score reduction in the HAM-D-17 total score is plotted in Figure 3A. The figures clearly demonstrate that early improvement is a necessary prerequisite of later stable response, as almost none of the patients who had *not* improved by week 2 later became a stable responder, indicating high sensitivity of early improvement as a predictor of stable response. On the other hand, a substantial number of patients with early improvement did not become stable responders, which expresses limited specificity of the predictor.

When the response criterion was chosen even more strictly by analyzing *stable remission* (defined as displaying HAM-D-17 total score of 7 or less at weeks 4 and 6), we found that the sensitivity of the predictor "early improvement" was even more pronounced (week 1: mirtazapine, 88%; paroxetine, 67%; week 2: mirtazapine, 100%; paroxetine, 92%; from week 3 onward both compounds had a sensitivity of 100%). Similarly, the NPVs were also higher than for the stable response criterion (week 1: mirtazapine, 92%; paroxetine, 86%; week 2: mirtazapine, 100%; paroxetine, 95%; from week 3 onward both compounds had NPVs of 100%).

The number of patients with available data sets from the ITT sample who achieved stable remission (HAM-D-17 ≤ 7 points at weeks 4 and 6) with or without improvement of at least 20% reduction in the HAM-D-17 total score is plotted in Figure 3B. These figures clearly demonstrate that almost none of the patients who had *not* improved by week 2 later became a stable remitter, indicating that early improvement is a necessary prerequisite of later stable remission.

To give an estimate of the suitability of early improvement to predict stable response, we calculated the areas under the ROC curves (AUC) for both drugs tested. While for this calculation the lower threshold value, implying prediction of stable response by chance, would be 0.5,

Figure 3A–B. Numbers of Patients in the Mirtazapine (MIR) and Paroxetine (PAR) Treatment Groups Who Were Found to Fulfill the Criterion of Improvement at Week X and Later Became Stable Responders (A) or Stable Remitters (B)[a]



[a]Improvement, response, and remission defined by scores on the 17-item Hamilton Rating Scale for Depression.

the values for mirtazapine and paroxetine were 0.81 and 0.73, respectively, indicating good predictability for both compounds.

DISCUSSION

Our analyses lead to the following conclusions:

1. Improvement as defined as a score reduction of at least 20% of the HAM-D-17 total score was observable in a majority of patients within the first 2 weeks of treatment with both drugs. This finding is consistent with the results of Stassen and colleagues,[8,10] who postulated that improvement is reliably detectable in the early phase of treatment in patients with major depression treated with antidepressants. The criterion of ≥ 20% score reduction has been chosen in order to allow a direct comparison to the work of Stassen and coworkers. The rationale to choose this cut-off value has been described by Stassen as follows: the random fluctuations of the HAM-D-17 total score in clinical trials have been empirically found to lie around 15%. Therefore, a value of 20% score reduction exceeds the random fluctuations caused by, for example, rater variability. Moreover, from a clinical point of view, this chosen cut-off represents a meaningful and reliably observable change in the clinical status of the patient. If a moderately depressed patient had an initial total HAM-D-17 score of 20 points, a 20% score reduction would mean a decrease of 4 points on the HAM-D.[19] This reduction has been shown to be reliably measurable.[14] In more severely depressed patients, the score reduction would be even more pronounced.
2. In our sample of patients, early improvement (at least 20% score reduction of HAM-D-17) predicted later stable response with remarkable high sensitivity. This is the most important finding from our report. After week 2, the sensitivity to predict later stable response on the basis of early improvement was as high as 97% for mirtazapine, but also with paroxetine (91%) a remarkably high sensitivity was calculated. In the context of predicting later stable response by a predictor (early improvement), a high sensitivity and low percentages of false negatives are desired (i.e., testing negative for the predictor makes a later response unlikely). Translated into clinical terms, this means that as long as the clinician observes improvement, he can continue with the treatment strategy. However, not all improvers will become responders (which is expressed by the limited specificity or the rate of false positives). On the other hand, if improvement is not observed, a high sensitivity indicates that a later stable response will be unlikely. According to our data, a clinician can decide by week 2 to 3 in patients not improving to change the treatment strategy, which is remarkably early in the course of treatment.
3. In a majority of patients, improvement, once it was observed, was subsequently maintained. This argues against the notion that improvement is only a stochastic phenomenon occurring in the course of treatment.
4. Choosing different cut-offs for the definition of improvement (i.e., at least 25% or 30% improvement) yielded comparable results. With higher cut-off thresholds, the sensitivity tended to decrease, while the specificity increased. Because a high sensitivity of the predictor is desired for clinical reasons, the 20% criterion seems most appropriate.
5. Early improvement was a highly sensitive predictor not only of stable response, but also of stable remission, which is clinically even more important. From our data, it can clearly be derived that early improvement within the first 2 weeks of treatment is a necessary prerequisite of later stable remission.

Taken together, our results argue against the hypothesis that antidepressants have a delayed onset of action

and confirm previous reports of Stassen et al., who reported for different types of antidepressants that early improvement predicted later response. However, the methodology applied in the work by Stassen et al. was based on survival analytical techniques, whereas we aimed to calculate sensitivity and specificity of the predictor "early improvement" for the criteria "stable response" and "stable remission," the latter representing a more rigorous clinical definition in our approach.

Quitkin and coworkers also stated that "early improvement is associated with a favorable prognosis."[5(p393)] However, as they focus in their work on placebo-drug differences, they concluded that the same data "do not support the hypothesis that antidepressant effect can be observed within the first 2 weeks of initiating treatment," because "ultimate responders" to both placebo and active drug did not significantly differ in the rates of early improvers.[5(p393)] Exactly in this point they confirm the work of Stassen, who postulated that placebo and active drug show similar time patterns of response in those who benefit from treatment. The difference between placebo and active drug lies only in the proportion of patients who show the pattern of response, but the pattern is similar for placebo and active drug.[21]

For the individual patient starting a treatment with a given drug, the problem of placebo-drug difference is only secondary. Of major interest, however, is the question, whether a later stable response or remission can be predicted early in the course of treatment, once the individual treatment is started. We have demonstrated in this report that early improvement is a highly sensitive predictor of later stable response and stable remission. Thus, our data are compatible with the results published by Quitkin and coworkers and extend the findings of Stassen and colleagues in a practically relevant manner.

We consider several caveats in the interpretation of our results. First, they are confined to a sample of mainly moderately depressed outpatients, who were neither chronically depressed nor treatment-resistant nor had any relevant anxiety disorder or a psychotic or bipolar illness history. Nevertheless, our sample represents a group of patients often encountered in clinical practice (especially in primary care), and at least for these patients the conclusions seem fully appropriate.

Second, our patients have not been treated with additional psychotropic drugs like benzodiazepines, which may influence the HAM-D-17 scores substantially.[20] It is not known whether the results reported here can be generalized to comedicated patients.

Third, so far the results are confined only to the 2 antidepressants tested. Prediction analyses have to be carried out for other agents in order to allow us to generalize the results presented here.

Fourth, we have not examined a placebo-treated group in this trial. However, in the work of Stassen et al., the pattern of response to placebo was not different from effective antidepressants.[21] The main difference between placebo and active antidepressants was found to lie in the number of patients showing improvement and subsequent response, but not in the time point at which improvement and response occurred. This finding was interpreted to suggest that antidepressants may act as "kick-starters" of the pattern of response observed under placebo.[21]

Fifth, we underscore that this was a rather short treatment trial, and one does not know whether the findings would change substantially if this were a treatment trial that lasted for 8 or 12 weeks.

There is growing evidence that antidepressant response starts during the early phase of treatment. Nierenberg and colleagues[9] recently reported that more than half of the eventual responders under fluoxetine started to respond by week 2. Furthermore, the same group found in a previous report that early nonresponse to fluoxetine predicted poor 8-week outcome.[22] Together with our data and the reports of Stassen et al., these reports imply important clinical consequences. It is recommended for the clinician to evaluate the depressive symptomatology of a given patient with the HAM-D-17 scale at baseline and in weekly intervals. Importantly, the clinician should then focus on the presence or absence of improvement (≥ 20% score reduction) in the early course of treatment. If improvement is observed, treatment should be continued. If improvement is not observed, reasonable decisions about the change of treatment strategies can be made much earlier than currently believed by most clinicians. In this regard, our results may help to reduce the time of suffering for patients, save time, and reduce treatment costs.

Drug names: clonidine (Catapres and others), fluoxetine (Prozac and others), mirtazapine (Remeron), paroxetine (Paxil), prazosin (Minipress and others).

REFERENCES

1. Everitt BS. Analysis of longitudinal data: beyond MANOVA. Br J Psychiatry 1998;172:7–10
2. Montgomery SA. Are 2-week trials sufficient to indicate efficacy? Psychopharmacol Bull 1995;31:41–44
3. Quitkin FM, Rabkin JD, Markowitz JM, et al. Use of pattern analysis to identify true drug response: a replication. Arch Gen Psychiatry 1987;44:259–264
4. Quitkin FM, Stewart JW, McGrath PJ, et al. Further evidence that a placebo response to antidepressants can be identified. Am J Psychiatry 1993;150:566–570
5. Quitkin FM, McGrath PJ, Stewart JW, et al. Can the effects of antidepressants be observed in the first 2 weeks of treatment? Neuropsychopharmacology 1996;15:390–394
6. Khan A, Cohen S, Dager S, et al. Onset of response in relation to outcome in depressed outpatients with placebo and imipramine. J Affect Disord 1989;17:33–38
7. Katz MM, Koslow SH, Frazer A. Onset of antidepressant activity: reexamining the structure of depression and multiple actions of drugs. Depress Anxiety 1996;4:257–267
8. Stassen HH, Delini Stula A, Angst J. Time course of improvement under antidepressant treatment: a survival-analytical approach [published erratum appears in Eur Neuropsychopharmacol 1993;3:543]. Eur

Neuropsychopharmacol 1993;3:127–135
9. Nierenberg AA, Farabaugh AH, Alpert JE, et al. Timing of onset of antidepressant response with fluoxetine treatment. Am J Psychiatry 2000;157:1423–1428
10. Stassen HH, Angst J, Delini Stula A. Delayed onset of action of antidepressant drugs? survey of results of Zurich meta-analyses. Pharmacopsychiatry 1996;29:87–96
11. Stassen HH, Angst J, Delini Stula A. Fluoxetine versus moclobemide: cross-comparison between the time courses of improvement. Pharmacopsychiatry 1999;32:56–60
12. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(suppl 20):22–33; interview 34–57
13. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC: American Psychiatric Association; 1994
14. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;6:278–296
15. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382–389
16. Guy W. ECDEU Assessment Manual for Psychopharmacology. US Dept Health, Education, and Welfare publication (ADM) 76-338. Rockville, Md: National Institute of Mental Health; 1976:218–222
17. Dunn G. Statistics in Psychiatry. London, England: Oxford University Press; 2000
18. Benkert O, Szegedi A, Kohnen R. Mirtazapine compared with paroxetine in major depression. J Clin Psychiatry 2000;61:656–663
19. Collegium Internationale Psychiatriae Scalarum (CIPS). Internationale Skalen für Psychiatrie. Göttingen, Germany: Beltz Test; 1996
20. Laakmann G, Faltermaier-Temizel M, Bossert-Zaudig S, et al. Treatment of depressive outpatients with lorazepam, alprazolam, amitriptyline and placebo. Psychopharmacology (Berl) 1995;120:109–115
21. Stassen H, Angst J. Delayed onset of action of antidepressants: fact or fiction? CNS Drugs 1998;9:177–184
22. Nierenberg AA, McLean NE, Alpert JE, et al. Early nonresponse to fluoxetine as a predictor of poor 8-week outcome. Am J Psychiatry 1995;152:1500–1503

