Is There a Common Resilience Mechanism Underlying Antidepressant Drug Response? Evidence From 2848 Patients

Hans H. Stassen, Ph.D.; Jules Angst, M.D.; Daniel Hell, M.D.; Christian Scharfetter, M.D.; and Armin Szegedi, M.D.

Objectives: Timing issues of antidepressant drug response are of major clinical relevance, given our current inability to predict when a particular patient will respond to a particular treatment.

Method: We detailed the time characteristics of recovery in a study of 2848 patients (diagnosed according to DSM-III-R/DSM-IV criteria as having major depressive disorder or major depressive episode) who were treated with 7 different antidepressants and placebo. A 2-dimensional cure model was used to disentangle the 2 central aspects of psychotropic drug response: the proportion of patients in whom a therapeutic response is induced (*incidence*) and the time to onset of improvement (*latency*). Random-effects models were applied to quantify unexplained heterogeneity. Patients were recruited between June 1982 and May 1998.

Results: Our analyses yielded no indication for a delayed onset of antidepressant drug response. Rather, we found highly individual time characteristics of recovery along with a continuous distribution of the time spans to onset of improvement under treatment with all active compounds and placebo. The mean ± SD time to onset of improvement was 13 ± 1 days and to response was 19 ± 1 days. Effective antidepressants appeared to trigger and maintain conditions necessary for recovery from the disorder. Odds-ratio analysis based on a random-effects model revealed that early improvers were at least 3 times more likely to become sustained responders with a pooled OR of 9.25, 95% CI = 7.79 to 10.98.

Conclusions: Affectively ill patients are likely to possess a common, biological, "resilience"-like component that largely controls recovery from depression. Once triggered, recovery appears to follow a pattern similar to the course observed with placebo, despite marked pharmacologic differences of the triggers. These findings may pave the way for new classes of psychotropic drugs specifically designed to support health-oriented processes underlying the natural resilience of patients.

(J Clin Psychiatry 2007;68:1195–1205)

Received July 14, 2006; accepted Dec. 20, 2006. From the Psychiatric University Hospital, Zurich, Switzerland (Drs. Stassen, Angst, Hell, and Scharfetter), and Organon International, Roseland, N.J. (Dr. Szegedi).

This project was funded in part by grants of the Swiss National Science Foundation (SNF32-42541.94, SNF31-63769.00).

Dr. Angst was a member of the speakers board for Roche, is currently a speaker for and advisor to Eli Lilly and AstraZeneca, and has received honoraria from Eli Lilly, GlaxoSmithKline, and Pfizer. Dr. Szegedi is an employee of Organon International. Drs. Stassen, Hell, and Scharfetter report no additional financial or other relationships relevant to the subject of this article.

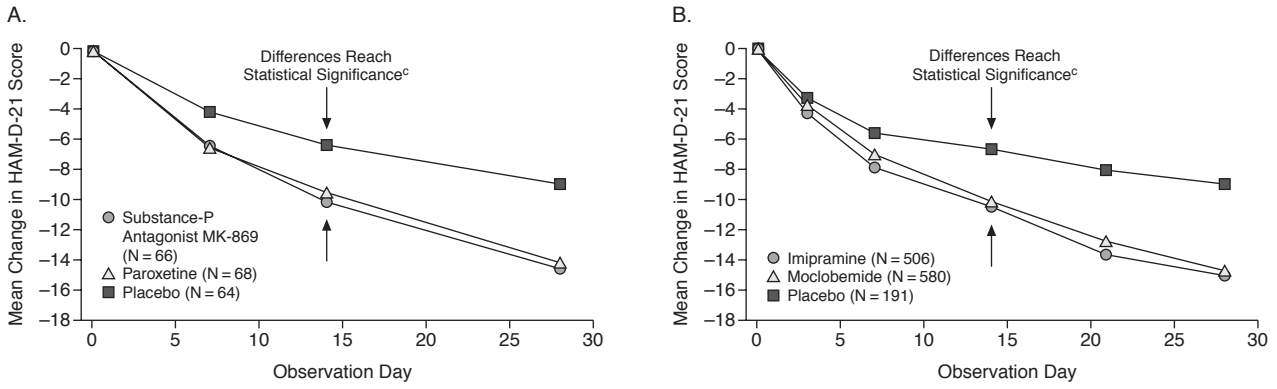
Corresponding author and reprints: Hans H. Stassen, Ph.D., Psychiatric University Hospital Zurich, P.O. Box 1931, CH-8032 Zurich, Switzerland (e-mail: k454910@bli.unizh.ch).

Current knowledge about the mechanisms by which antidepressants finally exert their therapeutic effects is rather limited despite decades of intensive research. In consequence, it is currently impossible to make any predictions of whether or not a particular patient will respond to a particular treatment. The most puzzling point in the treatment of major depressive disorder (MDD), however, is the observation that antidepressants, which differ greatly in their biochemical design and primary site of pharmacologic action, display virtually the same efficacy, as measured by the proportion of patients in whom they induce a therapeutic response.[1]

Assessing Efficacy of Antidepressants: The Average–Time-Course Approach

In general, results from placebo-controlled, double-blind studies yield comparable average time courses of improvement under treatment with effective compounds, along with very similar drug-placebo differences. Figure 1 reveals the striking similarity in mean change in the 21-item Hamilton Rating Scale for Depression (HAM-D-21) score under treatment with 4 pharmacologically different compounds compared to placebo: the substance-P antagonist MK-869 versus the selective serotonin reuptake inhibitor (SSRI) paroxetine (Figure 1A), and the tricyclic antidepressant imipramine versus the reversible inhibitor of monoamine oxidase-A (MAOA) moclobemide (Figure 1B). It is particularly worth noting that the differences between active compounds and placebo—as assessed for treatment groups by quantitative measures, such as the

Figure 1. Effect of Treatment on Mean Change From Baseline on the 21-Item Hamilton Rating Scale for Depression (HAM-D-21) for Patients With Major Depressive Disorder: (A) Substance-P Antagonist MK-869 or Paroxetine Versus Placebo[a] and (B) Imipramine or Moclobemide Versus Placebo[b]



[a]Data are from Kramer et al.[2]
[b]Data are from Stassen et al.[3]
[c]The time point at which differences reach statistical significance depends on sample size.

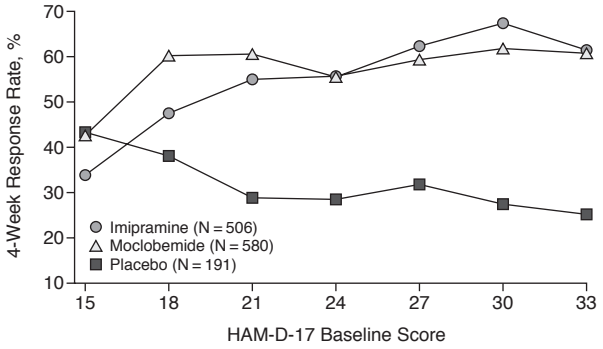
HAM-D rating scale—evolve over time in a very similar way from the very beginning of the drug trial, achieving over 65% of the final drug-placebo difference within the first 2 weeks of treatment,[4] when drug-placebo differences reach statistical significance. This statistical time lag has long been misinterpreted as indicating a delayed onset of action of antidepressants.

These finding and their misinterpretation are in line with recent results from 7450 patients treated with various types of antipsychotics, in which the greatest reduction in total scores on the Brief Psychiatric Rating Scale and the Positive and Negative Syndrome Scale occurred in the first week, with continuing drops in scores over subsequent weeks until a plateau was reached.[5]

Yet in addressing the question of antidepressant drug response, the average–time-course approach used for standard analyses of clinical drug trials is actually misleading, since detailed analyses of individual time courses of improvement readily reveal that averaging time courses across patients is too simplistic to grasp the complex processes underlying antidepressant drug response.[6] First, the efficacy of antidepressants in the treatment of MDD depends essentially on the severity of depression at baseline (Figure 2), with drug-placebo differences hardly ever reaching significance in mild MDD cases. For this reason, clinical drug trials typically require a minimum baseline score of at least 15 points on the 17-item HAM-D scale at entry into study (after washout), whereas Figure 2 suggests a minimum baseline score greater than 18 as inclusion criterion.

Second, scatterplots of individual depression scores versus time indicate considerable heterogeneity in the time courses of improvement (Figures 3A and 3B), which include early improvement, late improvement, partial improvement, nonimprovement, and premature withdrawal, the latter typically comprising 20% to 35% of the total sample.[8]

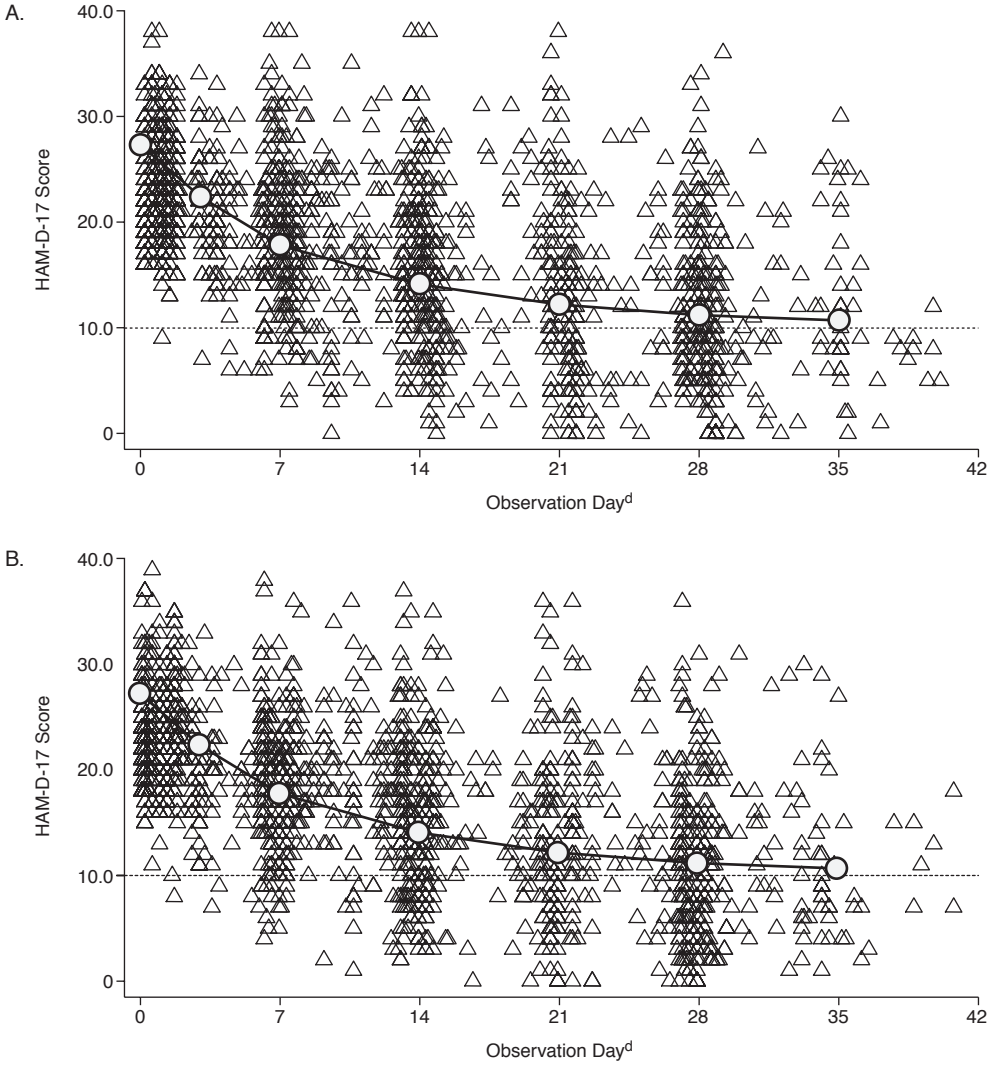
Figure 2. Response to Treatment With Imipramine, Moclobemide, and Placebo as a Function of the 17-Item Hamilton Rating Scale for Depression (HAM-D-17) Baseline Score After Placebo Washout[a]



[a]The 4-week response rates (50% baseline score reduction) are plotted along the y-axis for baseline scores 15 (HAM-D-17 ≤ 15), 18 (15 < HAM-D-17 ≤ 18), 21 (18 < HAM-D-17 ≤ 21) …, 33 (30 < HAM-D-17), indicating that placebo-drug differences reach 30% for HAM-D-17 baseline scores above 18.[7]

Also, these scatterplots reveal that the actual rating dates vary by ± 2 days around the prespecified "design" days of the standard study protocol (traditionally at weekly intervals), a variation that is due to the weekend organization of wards and clinical practices. All in all, the average–time-course approach to the analysis of clinical drug trials involves a considerable loss of information as to the various patterns of recovery from depression, and might imply imposing structure on empirical data rather than finding empirical structure in the data.

Figure 3. Time Course of Improvement Under Treatment With (A) Fluoxetine (selective serotonin reuptake inhibitor) and (B) Moclobemide (monoamine oxidase-A)[a,b,c]



[a]Data are from Stassen et al.[10]
[b]Single-case analysis: 440 patients (A) and 437 patients (B) with a diagnosis of major depression.
[c]The empirical 17-item Hamilton Rating Scale for Depression (HAM-D-17) scores of the patients are plotted along the y-axis as a function of the observation time (x-axis), which relates to the actual rating date rather than the ideal design days of the study protocol. The solid line denotes the underlying regression curve representing the average time course of improvement, which displays exponential characteristics, while the dotted line indicates the depression score of HAM-D-17 = 10, which is generally regarded as the threshold of "remission."
[d]To avoid overprinting in the scatter plots we added a negligible random variation d, $|d| \leq 0.5$, to the observation day.

Assessing Efficacy of Antidepressants: The Individual-Case Approach

Assessing the time course of improvement under treatment with antidepressants separately for each individual patient has several methodological advantages over the average–time-course model. In particular, the individual-case approach offers a way to disentangle the inter-individual diversity of antidepressant drug response, as well as to assess the speed of response under the various treatments at a much better resolution. Indeed, still today clinicians are puzzled trying to explain why (1) some MDD patients show a quick response to treatment (within a few days) with rapid sustained improvement and subsequent remission, while others experience a delayed onset of action; (2) some patients exhibit an irregular, fluctuating course of recovery, in which improvement may even get stuck at some point; and (3) a considerable proportion of MDD patients show no improvement at all under various kinds of treatment.

Therefore, the detection of natural groupings within a sufficiently representative sample of MDD patients could greatly facilitate the development of new, more specifi-

cally acting substances, hopefully without the side effects of today's drugs. Such developments are even more important given that current hypotheses on the primary mechanisms of action of antidepressants explain, if at all, only a small proportion of the observed inter-individual variation in response. A typical example of an unsuccessful attempt to directly link the effect of an antidepressant to its primary site of action relates to the MAOA hypothesis of moclobemide, in relation to which a molecular-genetic study yielded no significant relation to the respective genes.[9]

METHOD

The Zurich Study: Onset of Action of Antidepressants in 2848 MDD Patients

Using the individual-case approach, we investigated the time course of improvement under treatment with antidepressants in a sample of 2848 MDD patients (diagnosed according to DSM-III-R/DSM-IV criteria as having MDD or major depressive episode) originating from 4 independent drug trials and treated with 1 of 7 different types of antidepressants or placebo. The 4 drug trials were approved by institutional review boards. All patients of our sample (recruited between June 1982 and May 1998) had given their written informed consent and participated in double-blind clinical studies carried out to measure the efficacy of active compounds versus active comparators and/or placebo subsequent to a 3- to 7-day washout. The clinical data were computerized in our lab and cleaned according to Zurich standards. Details on the data set are given elsewhere.[5,10,11]

Specifically, we determined the time point at which each individual patient had achieved a certain HAM-D baseline score reduction, with time measures being based on the true assessment dates rather than the envisaged design days of the study protocols, thus enhancing the time resolution. As to the improvement criterion, we distinguished between "improvement" and "response" (20% and 50% baseline score reduction, respectively) and, when assessing the time point of onset of improvement, between unconstrained and sustained criteria, the latter requiring no subsequent deterioration reversing achieved improvement scores. Results based on unconstrained or sustained criteria differ in so far as more patients meet the unconstrained criteria along with an earlier onset of improvement. For prospective studies, unconstrained improvement criteria are mandatory.

Cure Models

The 2 central aspects of psychotropic drug response, the proportion of patients in whom a therapeutic response is induced (*incidence*) and the time to onset of improvement (*latency*), are often combined into a single, 1-dimensional "cure" model by means of survival-analytic techniques.[12] This cure model incorporates nonresponders in addition to modeling the time to an a priori defined clinical response. Patients who dropped out of the trial prior to the envisaged observation period are treated as censored data under the assumption of random censoring. The nonparametric generalized maximum likelihood for the model is obtained from the Kaplan-Meier product-limit estimator. By contrast, when separating incidence and latency through a 2-dimensional cure model, tests become available (Kolmogorov-Smirnov, Cramer-von Mises) to explicitly compare the speed of improvement *between treatments* among improvers.[13]

To detail the time characteristics of improvement, we calculated the cumulative rates of improvers and responders for each treatment modality as a function of observation day by means of the nonparametric Kaplan-Meier estimator.[14] Patients who did not improve were excluded because *un*conditional time to improvement curves combine (1) the proportion of patients who improve with (2) the time to improvement. Therefore, between-treatment comparison tests show statistical significance for differing proportions of improvers/responders even when the time characteristics of improvement are identical.[13]

Random-Effects Models

To ascertain differences between drug trials and between treatment modalities, we relied on random-effects models (generalized least squares models), which aim at quantifying unexplained heterogeneity in cases in which fixed-effects models seriously underestimate the standard errors for estimates and least squares means, thus biasing inferences.[15–18]

RESULTS

Early Improvement

Among patients who met the improvement or response criteria, the distribution of the individual onsets was left-skewed, yet approximately normal, covering a wide range from early to late improvement/response with mean values around day 12 to 14 (improvement) and day 18 to 20 (response). Unexpectedly, there were almost no differences between treatment modalities in this respect (Table 1), with differences between active compounds, and between active compounds and placebo, being reflected only by the total number of improvers and responders. The subtle differences in the average onset are likely the effect of nonstandardized placebo washouts, which may explain a ± 2-day shift of baseline, or may relate to the dose titration differences in the first days of treatment.

Analyses carried out separately for male and female patients did not reveal major gender differences, except for mirtazapine and paroxetine studies, in which sample sizes were small (Table 2). Meeting the improvement criteria was highly predictive of later outcome, as, typically,

Table 1. Mean Time to Sustained Response to Therapy[a,b,c]

		Patients Meeting 20% Criterion		Patients Meeting 50% Criterion	
Treatment	N	Proportion, (N) %	Day, Mean ± SD	Proportion, (N) %	Day, Mean ± SD
Imipramine	509	(420) 82.5	10.6 ± 7.2	(311) 61.1	18.0 ± 7.9
Moclobemide	1020	(801) 78.5	12.5 ± 8.1	(547) 53.6	18.9 ± 8.3
Placebo	316	(184) 58.2	13.3 ± 10.2	(100) 31.6	20.0 ± 10.0
Amitriptyline	167	(134) 80.2	12.2 ± 9.4	(96) 57.5	21.3 ± 11.4
Fluoxetine	444	(341) 76.8	14.2 ± 8.4	(211) 47.5	20.2 ± 8.3
Oxaprotiline	120	(77) 64.2	15.7 ± 11.2	(47) 39.2	23.3 ± 11.6
Mirtazapine	138	(111) 80.4	14.4 ± 11.4	(78) 56.5	19.4 ± 12.2
Paroxetine	134	(102) 76.1	14.2 ± 11.5	(71) 53.0	23.9 ± 11.6

[a]Derived/computed from data of Stassen et al.[3,10] and Szegedi et al.[11]
[b]Time points of sustained improvement (20% criterion) and sustained response (50% criterion) in the "average" patient under antidepressant and placebo treatment, as derived from an individual-case analysis using true assessment days rather than the design days of the study protocols.
[c]Differences between treatment modalities appear to relate to the proportion of patients in whom a therapeutic effect is induced but not to the onset of effect.

Table 2. Proportion of Sustained Improvers/Responders[a,b]

	Male Patients			Female Patients		
Treatment	N	20% Criterion, (N) %	50% Criterion, (N) %	N	20% Criterion, (N) %	50% Criterion, (N) %
Imipramine	146	(122) 83.6	(93) 63.7	363	(298) 82.1	(217) 59.8
Moclobemide	346	(272) 78.6	(181) 52.3	674	(529) 78.5	(361) 53.6
Placebo	98	(63) 64.3	(31) 31.6	218	(121) 55.5	(68) 31.2
Amitriptyline	54	(43) 79.6	(34) 63.0	113	(91) 80.5	(62) 54.9
Fluoxetine	156	(124) 79.5	(72) 46.2	288	(217) 75.3	(139) 48.3
Oxaprotiline	38	(23) 60.5	(16) 42.1	82	(54) 65.9	(31) 37.8
Mirtazapine	51	(40) 78.4	(23) 45.1	87	(71) 81.6	(55) 63.2
Paroxetine	46	(36) 78.3	(20) 43.5	88	(66) 75.0	(51) 58.0

[a]Derived/computed from data of Stassen et al.[3,10] and Szegedi et al.[11]
[b]Gender differences of sustained improvement (20% criterion) and sustained response (50% criterion) in the "average" patient under antidepressant and placebo treatment, as derived from an individual-case analysis using true assessment days rather than the design days of the study protocols.

greater than 70% of the patients who showed sustained improvement within the first 10 to 14 days of treatment (early improvement) later became sustained responders by the end of the 6-week observation period. Inversely, more than 80% of sustained responders at the end of the 6-week observation period had shown sustained improvement within the first 10 to 14 days of treatment, thus suggesting that in more than two thirds of the patients, the time course of improvement is sustained once the recovery has started. Allowing unconstrained improvement reduced the correct prediction of response by 3% to 5% because of additional patients with a fluctuating course of recovery who met the more permissive criteria.

Odds-ratio analysis based on a random-effects model (Figure 4) revealed that early improvers were at least 3 times more likely to become sustained responders with a pooled OR of 9.25 and 95% confidence interval = 7.79 to 10.98.

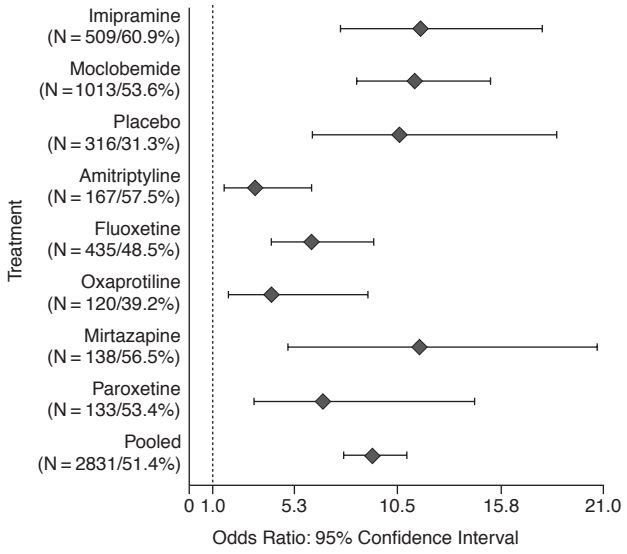
Incidence Versus Latency

Our results confirmed previous observations suggesting that the differences among active compounds, and between active compounds and placebo, manifest in the proportion of patients in whom an effect is induced (incidence), but not in the time characteristics of recovery (latency). Indeed, we find in Table 3 virtually the same cumulative rates of improvers over observation days, whereas the proportions of patients showing final improvement differ significantly between imipramine (83%), moclobemide (78%), and placebo (58%).

The same is true for the cumulative rates of improvers and responders under treatment with fluoxetine in comparison to moclobemide, in which no differences in the proportion of patients who showed improvement or response were found (Table 4). The apparent time lag of 3 to 4 days between the time scales underlying Tables 3 and 4 reflects current designs of clinical studies, with nonstandardized washouts and weekly assessments (plus additional assessments at days 3 and 10), which necessarily lack an absolute time scale, so that the question of the true mean ± SD time to onset of improvement (13 ± 1 days) or response (19 ± 1 days) remains open.

All in all, Tables 3 and 4 suggest that (1) the time course of improvement is independent of treatment modality; (2) effective antidepressants appear to trigger and maintain conditions necessary for recovery from the disorder; (3) once triggered, recovery appears to follow a pattern similar to the course observed with placebo, despite marked pharmacologic differences of the triggers; (4) approximately 95% of patients showing sustained

Figure 4. Odds Ratios Showing Increased Likelihood of Early Improvement Leading to Sustained Response[a,b,c,d]



[a]Sustained response is much more likely among early improvers independent of treatment modality as indicated by odds ratios that exceed 3 for all drugs and placebo.
[b]The estimates were derived by fitting a random-effects model that accounted for the between-study variation.
[c]The dotted line indicates an odds ratio of 1.
[d]Early improvement was defined through a 20% sustained baseline score reduction within the first 2 weeks of treatment and response through a 50% sustained baseline score reduction (criterion typically met around day 20 (mean 19.5 ± 9.2); cf. Table 1).

Table 3. Time Characteristics of Improvement Under Treatment With Imipramine, Moclobemide, and Placebo[a,b,c,d,e]

Day	Imipramine, n (N) = 419 (506),[f] %	Moclobemide, n (N) = 453 (580),[g] %	Placebo, n (N) = 111 (191),[h] %
3	15.4	17.8	23.3
7	55.1	53.7	54.3
10	61.9	61.2	60.3
14	81.3	80.1	77.6
21	93.5	92.1	89.7
28	97.7	97.0	95.7

[a]Data are from Stassen et al.[3]
[b]Cumulative rates of improvers under treatment with imipramine (TCA), moclobemide (MAOA), and placebo, as derived by survival-analytic methods.
[c]Onset of improvement was defined as a 20% Hamilton Rating Scale for Depression baseline score reduction without subsequent deterioration.
[d]Percentages relate to improvers only, excluding nonimprovers.
[e]The data of 3 patients were not yet available at the time point of the first analysis.
[f]Imipramine: 83% of total sample.
[g]Moclobemide: 78% of total sample.
[h]Placebo: 58% of total sample.
Abbreviations: MAOA = monoamine oxidase-A inhibitor, TCA = tricyclic antidepressant.

improvement, and greater than 85% of patients showing sustained response in a 6-week clinical trial, experience the respective onset within the first 4 weeks of treatment; and (5) there is no indication of a general 4-week delay of the onset of action of antidepressants, as proposed by some researchers in the field.[19]

DISCUSSION

Delayed Onset of Action of Antidepressants?

It is currently impossible to predict when a particular patient will respond to a particular antidepressant drug treatment—if at all. Moreover, response rates are generally modest,[20] as indicated by the observed drug-placebo and drug-drug differences in our study, in which drug effects were achieved through a variety of different mechanisms. This is in line with recent meta-analyses of U.S. Food and Drug Administration (FDA) data on 10,030 patients from 52 antidepressant drug trials, in which active substances showed superiority to placebo in fewer than half of the studies.[21,22]

Central to our study was the question of the extent to which active compounds possess a faster or slower onset of action, compared to placebo. To address this issue, we relied on a 2-dimensional cure model that disentangles the 2 central aspects of antidepressant drug response, efficacy in terms of the proportion of patients in whom a therapeutic response is induced (incidence), and efficacy in terms of the speed at which depressive symptoms reduce (latency).

We found a continuous distribution of the time spans to onset of improvement for all treatment modalities under investigation and for all improvement criteria on the basis of sustained baseline score reductions between 20% and 50%. The continuous distributions yielded no indication of a distinct drug effect after 3 to 4 weeks of treatment, as postulated by the delayed onset-of-action hypothesis.[19] That is, there was no indication of multimodal characteristics in which the first modal value is hypothesized to reflect placebo response and the second mode, after a delay of 3 to 4 weeks, is hypothesized to reflect true drug response. In fact, the delayed onset-of-action hypothesis has been seriously challenged by several recent meta-analyses of samples of respectable size.[1,17,23,24]

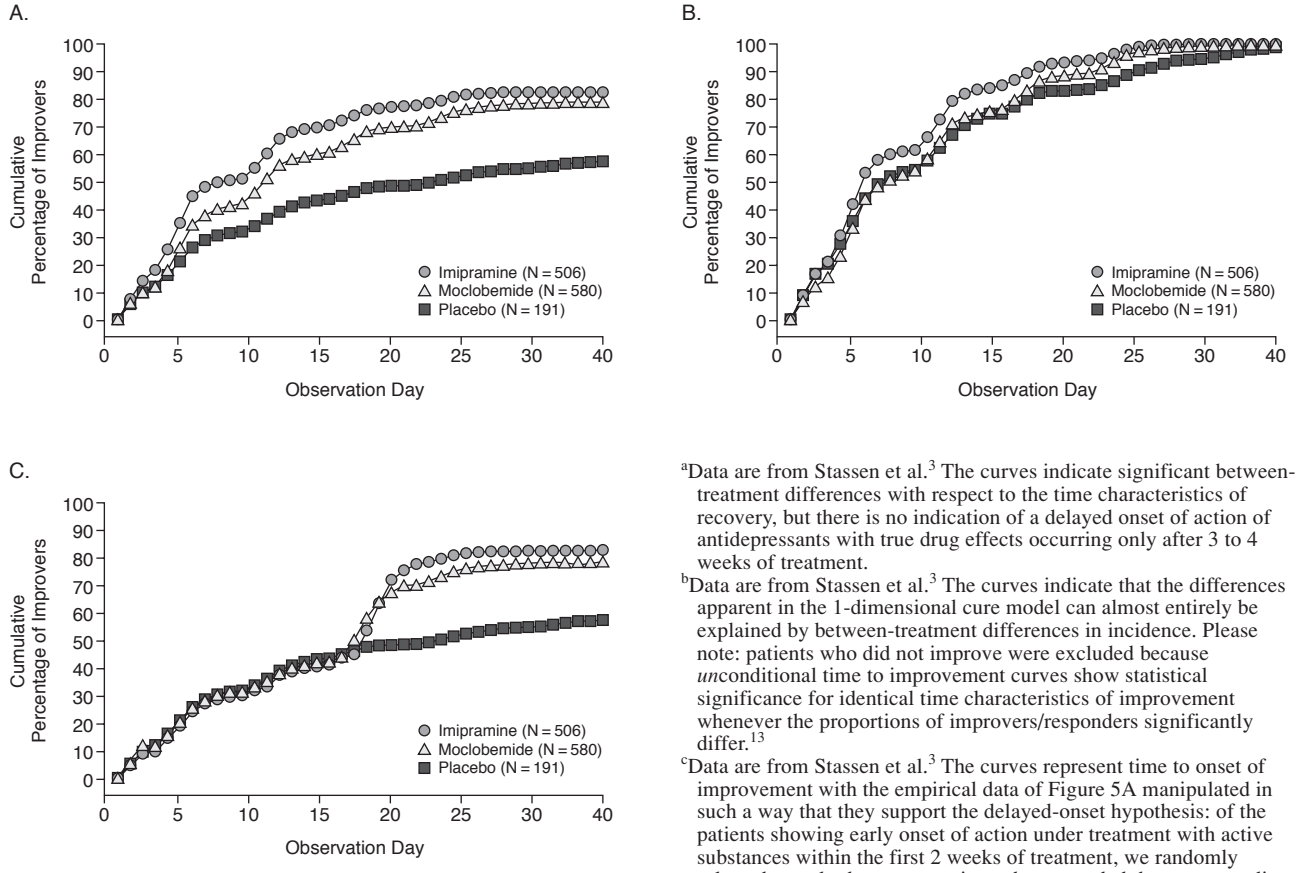
Specifically, the speed differences among active compounds and between active compounds and placebo turned out to be marginal, yet became statistically significant when incidence and latency were combined into a single cure model (Figure 5A), so that the time to improvement was weighted (biased) by the treatment-dependent response rates. It is likely, thus, that reports on "faster-acting drugs" can almost entirely be explained by the specifics of 1-dimensional cure models or equivalent statistical procedures when compared to the results of 2-dimensional models (Figure 5B). If one drug displays better efficacy than another, it apparently converts throughout the entire observation period a higher percentage of

Table 4. Time Characteristics of Improvement Under Treatment With Fluoxetine and Moclobemide[a,b,c,d,e]

	Criterion: 20% Sustained Baseline Score Reduction		Criterion: 50% Sustained Baseline Score Reduction	
Day	Fluoxetine, n (N) = 345 (440),[f] %	Moclobemide, n (N) = 348 (437),[g] %	Fluoxetine, n (N) = 211 (440),[h] %	Moclobemide, n (N) = 209 (437),[i] %
5	7.7	6.5	1.4	1.0
7	33.7	33.8	10.1	11.7
10	43.5	43.7	17.3	11.7
14	67.5	66.8	34.1	40.3
21	79.0	81.9	57.7	59.7
28	94.7	94.8	88.5	87.9

[a]Data are from Stassen et al.[10]
[b]Cumulative rates of improvers and responders under treatment with fluoxetine (SSRI) and moclobemide (MAOA) as derived by nonparametric survival-analytic methods.
[c]Onset of improvement and response were defined as a 20% and a 50% Hamilton Rating Scale for Depression baseline score reduction without subsequent deterioration, respectively.
[d]Percentages relate to improvers/responders only, excluding nonimprovers/nonresponders.
[e]The data of 7 patients (3 moclobemide, 4 fluoxetine) were not yet available at the time point of the first analysis.
[f]Fluoxetine: 78% of total sample were improvers.
[g]Moclobemide: 80% of total sample were improvers.
[h]Fluoxetine: 48% of total sample were responders.
[i]Moclobemide: 48% of total sample were responders.
Abbreviations: MAOA = monoamine oxidase-A inhibitor, SSRI = selective serotonin reuptake inhibitor.

Figure 5. Time to Onset of Improvement Under Treatment With Imipramine, Moclobemide, and Placebo: (A) One-Dimensional Cure Model Combining Incidence and Latency,[a] (B) Two-Dimensional Cure Model Separating Incidence and Latency,[b] and (C) One-Dimensional Cure Model as One Would Expect Under the Hypothesis of a Delayed Onset of Action of 3 Weeks[c]



[a]Data are from Stassen et al.[3] The curves indicate significant between-treatment differences with respect to the time characteristics of recovery, but there is no indication of a delayed onset of action of antidepressants with true drug effects occurring only after 3 to 4 weeks of treatment.
[b]Data are from Stassen et al.[3] The curves indicate that the differences apparent in the 1-dimensional cure model can almost entirely be explained by between-treatment differences in incidence. Please note: patients who did not improve were excluded because *un*conditional time to improvement curves show statistical significance for identical time characteristics of improvement whenever the proportions of improvers/responders significantly differ.[13]
[c]Data are from Stassen et al.[3] The curves represent time to onset of improvement with the empirical data of Figure 5A manipulated in such a way that they support the delayed-onset hypothesis: of the patients showing early onset of action under treatment with active substances within the first 2 weeks of treatment, we randomly selected exactly that many patients that exceeded the corresponding placebo rate and randomly shifted their onset of action between day 17 and day 24.

patients to responders who otherwise would remain nonresponders, a finding that is compatible with the results of Posternak and Zimmerman[4] on early drug-placebo separation. If there were a distinct drug effect only after 3 weeks of treatment, the curves derived from active substances and placebo should be identical throughout the first 3 weeks (Figure 5C).

In consequence, our results might explain the relative lack of progress in the field of psychotropic drug research to some extent, as the effects of the various classes of antidepressants—similar in magnitude, although achieved through different pharmacologic sites of action—appear to be indirect and unspecific with respect to the clinically defined target syndromes.[25] Indeed, many psychotropic drugs appear to be polyvalent in the sense that they induce a therapeutic response in various clinical indications. For example, antidepressants are also effective in various anxiety disorders or pain syndromes, and neuroleptics are successfully used in the treatment of bipolar illness,[26] among others. On the other hand, epidemiologic data suggest elevated comorbidity of MDD with other clinical syndromes, as is the case with coronary artery disease.[27,28]

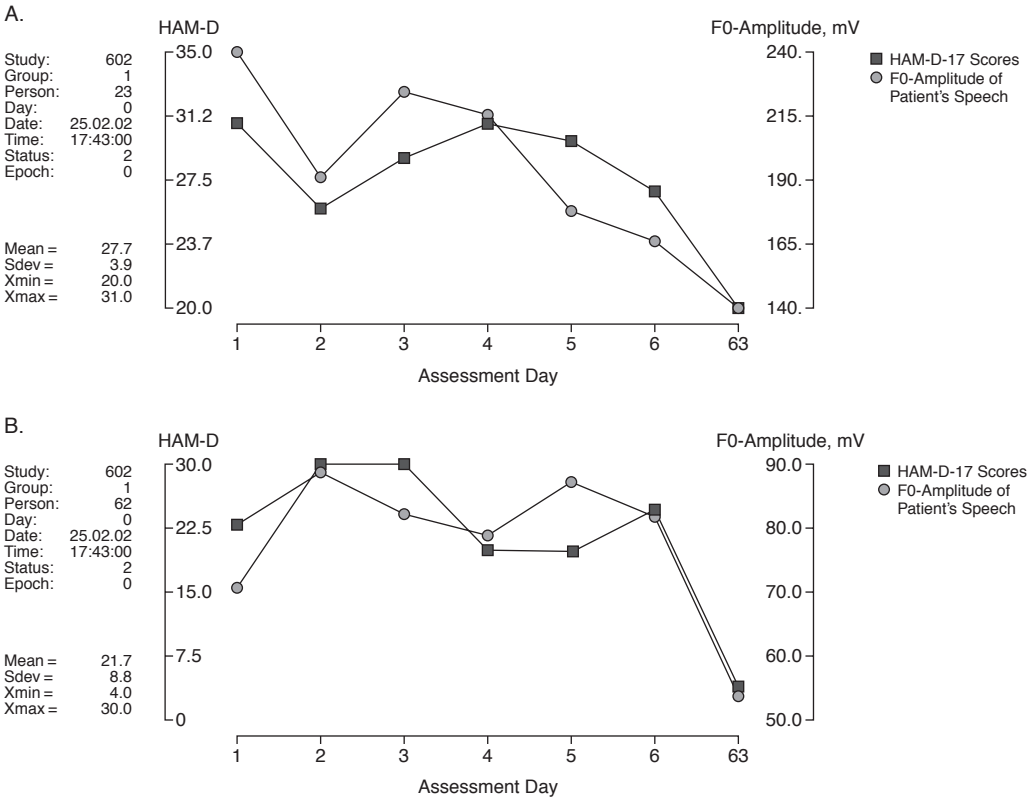
Undoubtedly, none of the studies conducted so far to investigate onset of action of antidepressants has been specifically designed for measuring time-effect differences among antidepressants or between antidepressants and placebo. Therefore, leading experts in the field have developed basic concepts along with an ideal study design for assessing differential onset of action in antidepressant efficacy trials when comparing 2 active drugs with placebo.[29] Apart from this ideal design, the authors underlined that more realistic studies would nonetheless provide valuable information on the timing issues of antidepressant drug response if certain limitations were taken into account. Specifically, onset of action should (1) be defined in a prospective way; (2) involve clinically relevant reduction in depressive symptoms; (3) be linked to clinically relevant treatment outcome; and (4) be analyzed not only for endpoint completers, but also for premature withdrawals and patients with missing data at some point of the trial.

To meet these criteria, we carried out a model-finding study by analyzing data from a 1-week placebo run-in, comprising 3 repeated HAM-D assessments. We found that the observed fluctuations did not exceed 15% of baseline score in the vast majority of cases over the 1-week observation period. In consequence, we defined *onset of improvement* (which models onset of action) as a 20% sustained baseline score reduction[14] in accordance with clinical practice in which a 4-point HAM-D reduction (= 20% for a HAM-D baseline score of 20) is regarded as clinically relevant. Similarly, we defined response as a sustained 50% baseline score reduction—generally regarded as a clinically relevant treatment outcome—and determined the interrelation between the so-defined onset of improvement and later response. It turned out that onset of improvement is highly predictive of later outcome (odds ratio of 9.25), thus underlining the potential value of the onset concept. It is the specific advantage of cure models over standard cross-sectional approaches that both (1) patients who prematurely withdraw from a clinical trial are kept in the analysis as censored data until the time point of dropout, thus avoiding biases inherent in the standard last-observation-carried-forward method; and (2) the variation of typically ± 2 days around the prespecified assessment days of the study protocol are taken into account, whereas this variation is ignored in cross-sectional analyses. From the scientific point of view, the onset concept directs attention to a key question of antidepressant drug research: Why do some 20% to 25% of patients who show an initial onset of action get stuck at some point of their recovery and eventually fail to meet response criteria? Identification of this subgroup of patients by operational criteria is one of the benefits of the early onset approach, and is likely to yield new insights into the mechanisms of action of antidepressants or psychotropic drugs in general.

Extending clinical trials beyond 6 weeks certainly identifies additional late responders (typically in the range of 5%) but does not yield additional or even principally new information about the time characteristics of recovery under treatment with antidepressants. This has been demonstrated, for example, by Thase and colleagues[30] in a meta-analysis comparing venlafaxine and SSRIs with placebo. For a variety of different response and remission criteria, the placebo-drug differences reached by the end of week 4 turned out to be similarly present by the end of week 8, and to accurately predict the latter.[30,31] Interestingly, there was an increase in placebo-drug difference from 72% to 80% between weeks 4 and 8, which almost exclusively originated from female patients.[31] Given these strong intrinsic regularities underlying the time course of recovery, we expect neither extending clinical trials beyond 6 weeks nor tightening up the stringency of response criteria to lead to principally new results as to the timing issues of antidepressant drug response.

To what extent do observer ratings, such as HAM-D or Montgomery-Asberg Depression Rating Scale, reflect the expectations of physicians and patients? We have addressed this question through objective laboratory methods and carried out a speech study on 63 depressed patients with 6 repeated assessments at 2-day intervals over the first 2 weeks plus a final assessment at the time point of hospital release (labeled "day 63" in Figure 6). Each assessment comprised a HAM-D rating at a fixed time in the morning plus a standardized voice recording of 8 minutes length. Speech recordings were taken in our speech laboratory by a technician.[32] In 65% of patients we found a close correlation over time ($r \approx 0.8$) between the

Figure 6. Time Course of Recovery From Depression of 2 Patients (A and B) as Reflected by 17-Item Hamilton Rating Scale for Depression Scores Versus the Speech Parameter F0-Amplitude[a,b]



[a]Data in part A are from Stassen et al.[32] Data in part B are unpublished (H.H.S.; S. Kuny, M.D.; D.H.; et al., 2002).
[b]Assessed at 2-day intervals over an observation period of 2 weeks, plus a final assessment at the time of discharge from hospital (day 63). The corresponding change over time of the speech parameter F0-amplitude (assessments taken while patients read out loud emotionally neutral text) is shown in order to demonstrate the close relationship between the 2 courses of development.

HAM-D scores and speech parameters (reference 32 and H.H.S.; S. Kuny, M.D.; D.H.; et al., unpublished data, 2002) (Figures 6A and 6B). Patients not showing such close correlations displayed either no change or an irregular pattern of nonimprovement.

Detailed data analysis suggested that observer ratings function well as long as there is clinical change. Score changes relative to baseline, if there are changes at all, appear to be reliably assessable by the majority of professional raters while the reliability of absolute scores is much weaker (less experienced raters tend to overrate). Therefore, rater training is an absolute must for multicenter clinical trials.[33] However, rater training does not necessarily solve the compatibility problems of absolute scores across centers. Moreover, extensive training is prone to overadaptation. The latter, in turn, increases the likelihood of imposing structure on data—rather than finding structure in the data. Given all these difficulties, one might argue that within-rater and between-rater variations obscure not only real differences among or between active compounds and placebo, but also differences in the time characteristics of recovery. This is certainly the case. On the other hand, differences in efficacy between active compounds or in their onset of action are generally modest and beyond major clinical relevance. It is quite unlikely that there exist strong and robust effects that are of major clinical relevance but systematically obscured by random noise in all studies worldwide—for example, that a 50% placebo-drug difference after 4 weeks or a shift toward earlier onset of greater than 1 week in a substantial proportion of patients[29] has gone undetected.

On the basis of FDA data, Khan and colleagues[34] recently compared placebo-drug differences observed in fixed-dose and flexible-dose efficacy trials. Symptom reduction was very similar under flexible- and fixed-dose treatment with active compounds (42.2%–42.9%), whereas the flexible- versus fixed-dose differences were much higher with placebo treatment: 29.3% symptom reduction with flexible dose (42.2% dropouts) versus 35.8% with fixed dose (35.3% dropouts), thus indicating a significantly better placebo response for the fixed-dose regimen. As a direct consequence, more placebo-drug dif-

ferences reached statistical significance in flexible-dose (58% for new antidepressants, 61.9% for active comparators) than in fixed-dose (30% for new antidepressants, 50% for active comparators) trials. The authors concluded that flexible dosing schedules in efficacy studies favor significant placebo-drug differences (provided the observed differences in symptom reduction with placebo are not attributable to the differences in dropout rates). Therefore, the question arises whether a flexible dose regimen can positively influence not only the proportion of patients in whom a therapeutic response is induced but also the time characteristics of recovery. Clinical experience clearly speaks in favor of flexible dose and rapid titration, but conclusive data are currently not available. Also, the extent to which less severe side effects that do not lead to premature withdrawal (e.g., minor weight gain) may influence treatment outcome is largely unknown.

Implications

If effective antidepressants merely trigger and maintain conditions necessary for recovery, as suggested by our results, then affectively ill patients are likely to have a biological, presumably genetically determined, predisposition that controls the time characteristics of recovery to a major extent and can be triggered, modulated, or suppressed by various endogenous and exogenous factors. The very similar, approximately normally distributed time points of onset of improvement—under all treatment modalities and with no gender differences—particularly support the existence of a genetic basis underlying these time characteristics. However, it appears quite unlikely that the respective distributions result from 1 single gene or a few major genes, as single gene approaches typically explain no more than a small percentage, less than 1.5%, of observed response-nonresponse dichotomy, and the etiologically heterogeneous group of nonresponders likely encompasses improvers, in whom onset of improvement has not yet been triggered, as well.

Oligogenic configurations of genes, in which nonlinear interactions can be larger than the main effects, and in which no single genomic locus is, by itself, either necessary or sufficient for the phenotype, may be more successful in modeling the complex phenomena of psychotropic drug response if combined with quantitative measures on the phenotype level.[35–37] The time characteristics of recovery—e.g., the times to onset of improvement—might well serve as such quantitative measures while, on the genotype level, receptors and transporters of serotonin, norepinephrine, dopamine, glutamate, glycine, γ-aminobutyric acid, acetylcholine, corticotropin-releasing hormone, and glucocorticoids; monoamine degrading enzymes such as monoamine oxidase and catechol *O*-methyltransferase; and metabolizing structures such as cytochrome P450 isoenzymes, along with various other transporters and transcription factors,[38] might constitute a "candidate configuration" of genes that can be tested for significant genotype-phenotype correlations. In fact, our recent pilot study of 257 patients and 178 candidate genes yielded highly significant correlations ($r \geq 0.45$; $p \leq .001$) between the time point of onset of improvement and a 23-dimensional genotype.

CONCLUSIONS

Our analyses yielded no indication for a delayed onset of antidepressant drug response. Rather, we found highly individual time characteristics of recovery along with a continuous distribution of the time spans to onset of improvement under treatment with all active compounds and placebo. The mean ± SD time to onset of improvement was 13 ± 1 days and to response was 19 ± 1 days. Effective antidepressants appeared to trigger and maintain conditions necessary for recovery from the disorder, with triggers being achieved at similar rates under treatment with substances that may differ greatly in their biochemical design and primary site of pharmacologic action. All this suggests that affectively ill patients may possess a biological, "resilience"-like component that controls recovery from depression to a major extent. The term *resilience* is used here as a broader concept than just neurogenesis, encompassing all those endogenous mechanisms that support and maintain health, thereby enabling patients to cope with stressful situations. This may include personality traits supporting or impeding social skills. Once triggered, recovery appears to follow a pattern similar to the course observed with placebo, despite marked pharmacologic differences of the triggers. Consequently, the vast majority of patients showing sustained response in a 6-week clinical trial experience the respective onset within the first weeks of treatment. These findings may clear the way to new classes of psychotropic drugs specifically designed to support the health-oriented processes underlying the resilience component of patients, thus increasing the number of patients who benefit from treatment, possibly in a prophylactic setting as well, and from reduced side effects. The latter prospect is of major clinical and economic relevance, given the large number of affectively ill patients.

Drug names: acetylcholine (Miochol-E), fluoxetine (Prozac and others), imipramine (Tofranil and others), mirtazapine (Remeron and others), paroxetine (Paxil, Pexeva, and others), venlafaxine (Effexor and others).

REFERENCES

1. Montgomery SA, Bech P, Blier P, et al. Selecting methodologies for the evaluation of differences in time to response between antidepressants. J Clin Psychiatry 2002;63:694–699
2. Kramer MS, Cutler N, Feighner J, et al. Distinct mechanism for antidepressant activity by blockade of central substance P receptors. Science 1998;281:1640–1645

3. 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
4. Posternak MA, Zimmerman M. Is there a delay in the antidepressant effect? a meta-analysis. J Clin Psychiatry 2005;66:148–158
5. Agid O, Kapur S, Arenovich T, et al. Delayed-onset hypothesis of antipsychotic action: a hypothesis tested and rejected. Arch Gen Psychiatry 2003;60:1228–1235
6. Holsboer F. Antidepressant drug discovery in the postgenomic era. World J Biol Psychiatry 2001;2:165–177
7. Angst J, Stassen HH. Methodische Aspekte von Studien zur antidepressiven Wirksamkeit. In: Steinberg R, Philipp M, Möller HJ, eds. Spezielle Aspekte der Antidepressiven Therapie: Neuere Ergebnisse zu Moclobemid. Munich, Germany: MMV Medizin Verlag München; 1994:147–165
8. Houck PR, Mazumdar S, Koru-Sengul T, et al. Estimating treatment effects from longitudinal clinical trial data with missing values: comparative analyses using different methods. Psychiatry Res 2004;129:209–215
9. Müller DJ, Schulze TG, Macciardi F, et al. Moclobemide response in depressed patients: association study with a functional polymorphism in the monoamine oxidase A promoter. Pharmacopsychiatry 2002;35: 157–158
10. Stassen HH, Angst J, Delini-Stula A. Fluoxetine versus moclobemide: cross-comparison between the time courses of improvement. Pharmacopsychiatry 1999;32:56–60
11. Szegedi A, Müller MJ, Anghelescu I, et al. Early improvement under mirtazapine and paroxetine predicts later stable response and remission with high sensitivity in patients with major depression. J Clin Psychiatry 2003;64:413–420
12. Laska EM, Meisner MJ. Nonparametric estimation and testing in a cure model. Biometrics 1992;48:1223–1234
13. Tamura RN, Faries DE, Feng J. Comparing time to onset of response in antidepressant clinical trials using the cure model and the Cramer-von Mises test. Stat Med 2000;19:2169–2184
14. Stassen HH, Delini-Stula A, Angst J. Time course of improvement under antidepressant treatment: a survival-analytical approach. Eur Neuropsychopharmacol 1993;3:127–135
15. Thase ME. Comparing the methods used to compare antidepressants. Psychopharmacol Bull 2002;36(suppl 1):suppl 1–17
16. Daniels MJ, Zhao YD. Modelling the random effects covariance matrix in longitudinal data. Stat Med 2003;22:1631–1647
17. Papakostas GI, Perlis RH, Scalia MJ, et al. A meta-analysis of early sustained response rates between antidepressants and placebo for the treatment of major depressive disorder. J Clin Psychopharmacol 2006;26: 56–60
18. Statistical analysis software SAS: Proc MIXED versus Proc GLM for random effects analysis. Cary, NC: SAS Institute Inc; 1999
19. Quitkin FM, Rabkin JD, Markowitz JM, et al. Use of pattern analysis to identify true drug response. Arch Gen Psychiatry 1987;44:259–264
20. Nemeroff CB, Thase ME, on behalf of the EPIC 014 Study Group. A double-blind, placebo-controlled comparison of venlafaxine and fluoxetine treatment in depressed outpatients. J Psychiatr Res 2007;41:351–359
21. Khan A, Khan SR, Leventhal RM, et al. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: a replication analysis of the Food and Drug Administration Database. Int J Neuropsychopharmacol 2001;4:113–118
22. Khan A, Khan SR, Leventhal RM, et al. Symptom reduction and suicide risk among patients treated with placebo in antipsychotic clinical trials: an analysis of the Food and Drug Administration database. Am J Psychiatry 2001;158:1449–1454
23. Angst J, Stassen HH. Do antidepressants really take several weeks to show effect? In: Leonard BE, ed. Antidepressants. Basel, Switzerland: Birkhäuser; 2001:21–30
24. Katz MM, Tekell JL, Bowden CL, et al. Onset and early behavioral effects of pharmacologically different antidepressants and placebo in depression. Neuropsychopharmacology 2004;29:566–579
25. Nierenberg AA, Wright EC. Evolution of remission as the new standard in the treatment of depression. J Clin Psychiatry 1999;60(suppl 22):7–11
26. McIntyre R, Katzman M. The role of atypical antipsychotics in bipolar depression and anxiety disorders. Bipolar Disord 2003;5(suppl 2):20–35
27. Rudisch B, Nemeroff CB. Epidemiology of comorbid coronary artery disease and depression. Biol Psychiatry 2003;54:227–240
28. Lett HS, Blumenthal JA, Babyak MA, et al. Depression as a risk factor for coronary artery disease: evidence, mechanisms, and treatment. Psychosom Med 2004;66:305–315
29. Leon AC, Blier P, Culpepper L, et al. An ideal trial to test differential onset of antidepressant effect. J Clin Psychiatry 2001;62(suppl 4):34–36
30. Thase ME, Entsuah AR, Rudolph RL. Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. Br J Psychiatry 2001;178:234–241
31. Entsuah AR, Huang H, Thase ME. Response and remission rates in different subpopulations with major depressive disorder administered venlafaxine, selective serotonin reuptake inhibitors, or placebo. J Clin Psychiatry 2001;62:869–877
32. Stassen HH, Kuny S, Hell D. The speech analysis approach to determining onset of improvement under antidepressants. Eur Neuropsychopharmacol 1998;8:303–310
33. Kobak KA, Lipsitz JD, Williams JB, et al. A new approach to rater training and certification in a multicenter clinical trial. J Clin Psychopharmacol 2005;25:407–412
34. Khan A, Khan SR, Walens G, et al. Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the food and drug administration summary basis of approval reports. Neuropsychopharmacology 2003;28:552–557
35. Burghes AH, Vaessin HE, de La Chapelle A. Genetics: the land between Mendelian and multifactorial inheritance. Science 2001;293:2213–2214
36. Stassen HH, Bridler R, Hell D, et al. Ethnicity-independent genetic basis of functional psychoses: a genotype-to-phenotype approach. Am J Med Genet B Neuropsychiatr Genet 2004;124:101–112
37. Henn FA, Vollmayr B. Neurogenesis and depression: etiology or epiphenomenon? Biol Psychiatry 2004;56:146–150
38. Taylor C, Fricker AD, Devi LA, et al. Mechanisms of action of antidepressants: from neurotransmitter systems to signaling pathways. Cell Signal 2005;17:549–557