A video method for the evaluation of antidepressant clinical and behavioural actions

Martin M. Katz, John P. Houston, Stephen Brannan, Janet Tekell, Nancy Berman, Charles L. Bowden and Alan Frazer

1 Department of Psychiatry, University of Texas Health Science Center, San Antonio, TX, USA
2 Eli Lilly and Company, Indianapolis, IN, USA
3 Cyberonics, Houston, TX, USA
4 Department of Psychiatry, VA Ann Arbor Healthcare System, University of Michigan, Ann Arbor, MI, USA
5 Harbor UCLA Medical Center, Torrance, CA, USA
6 Department of Pharmacology, University of Texas Health Science Center, San Antonio, TX, USA

Abstract

Measuring efficacy and behavioural actions of new antidepressants (ADs) is greatly enhanced by having videotaped records of assessment interviews. This study describes a revised version of the Video Interview Behaviour Evaluation Scales (VIBES), shortened to make it more applicable to clinical trials. The method focuses on physically expressive, motor and social behavioural aspects of the depressive disorder. The Brief version permits juxtaposing of baseline and outcome interviews during the same viewing session thereby reducing the role of memory in the rater’s observations. The method provides measures of behavioural components and four new severity dimensions, Social withdrawal–motor retardation, Anxiety–agitation, Hostility, Depressive mood–cognitive impairment. Viewing a patient’s series of baseline, during treatment, and outcome interviews can be conducted in about 1 h. This study compared the VIBES with the Hamilton Depression Rating Scale (HAMD) in (1) assessing the efficacy of the selectively targeted ADs, desipramine and paroxetine, and placebo in depressed in-patients and (2) determining onset and nature of the drugs’ early behavioural actions. The findings showed (1) components of the method to be reliable; (2) the VIBES to be more sensitive than the HAMD in measuring efficacy; (3) the methods to be equally sensitive in detecting early clinical actions of the two drugs; (4) the VIBES more informative in identifying discrete behavioural aspects of the disorder that are impacted by the drugs; and (5) that in differentiating the drugs’ behavioural actions, desipramine was indicated to initially ‘stimulate’, i.e. effect motor activity and depressed mood, and paroxetine to reduce global severity and anxiety. The study shows the VIBES to be capable of uncovering behavioural mechanisms underlying AD’s capacity to resolve depressive disorder.

Received 27 January 2005; Reviewed 10 March 2005; Revised 31 March 2005; Accepted 29 April 2005; First published online 22 July 2005

Key words: Antidepressants, behavioural actions, clinical trials, onset, video method.

Introduction

In a recent study, a set of established behavioural methods was used to determine onset and the differential nature of early clinical actions of a selective noradrenergic and a serotonergic antidepressant (AD) (Katz et al., 2004b). As part of the methodology, the Video Interview Behavioural Evaluation Scales (VIBES), aimed at elaborating the symptomatic, physical expressive and social behavioural actions of the drugs was also applied. Experience had shown video evaluation to have certain advantages over judgements from ‘live’ interviews, e.g. enhancing observation by permitting detachment from the interview, providing a focus on social behaviour and expressivity, and making possible the juxtaposing of baseline and outcome interviews, thereby reducing the role of memory in the rater’s judgement. The VIBES although successful in prior research in elaborating on the effects of a range of drugs, including the tricyclic antidepressants (Katz and Itil, 1974, Katz et al., 1989), has been difficult to harness for clinical studies, due both to technical problems and the tedium of
rating long interviews. To make it more applicable to clinical trials, the VIBES standardized interview and rating scales were substantially revised and markedly reduced in length. The method now provides measures of specific behavioural components and a set of severity dimensions. It permits ratings of an entire series of video interviews of the patient, baseline, during treatment, and at outcome, to be conducted in ~1 h.

Goals of this substudy were to test the revised video method (1) in its capacity to evaluate the early actions of the drugs on the major behavioural components of the depressive disorders; and (2) to determine its utility in a clinical trial of the efficacy of AD drugs. Such findings could also lead to identifying potential applications of the drug for use in treating other disorders. In this study, the derivation of the measures and their reliabilities are described, and their sensitivities are compared with that of an established measure, the Hamilton Depression Rating Scale (HAM-D; Hamilton, 1960).

The method is shown to be reliable, to elaborate significantly on the nature of the behavioural actions of the two drugs, and to be an effective measure of AD efficacy and the onset of clinical action.

Method

Research design

The study design was a randomized, parallel group, double-blind comparison of patients taking desipramine (DMI), paroxetine, or placebo for a period of 6 wk.

Patient sample

Patients with a diagnosis of primary major depression, unipolar type, single, or recurrent episode were identified from newly admitted in-patients at two Texas Veterans Administration (VA) hospitals. All subjects provided written informed consent and the study was carried out as approved by the University of Texas Health Center at San Antonio’s Institutional Review Board (IRB) and the Dallas VA Medical Center’s IRB. Diagnostic interviews were conducted using the Structured Clinical Interview for DSM-III (SCID; Spitzer and Williams, 1983). Patients were required to score ≥18 on the HAMD (21-item version) (Hamilton, 1960).

A total of 82 patients were initially enrolled and randomly assigned to treatment. Of these, 12 dropped out prior to receiving the minimum of 3 wk of treatment to qualify for the study. Reasons for dropping out are detailed in Katz et al. (2004b), as is further information on the clinical status of this depressed in-patient sample. The patients were, in accord with the CGI (Guy, 1976) ‘moderately’ to ‘severely’ depressed, the average patient score on the HAMD at baseline for the sample was 23.5 + 4.8 (± S.D.) and was essentially equivalent at the two sites. Of the 70 evaluable patients in the study proper (Katz et al., 2004b), videotaping of baseline, during treatment and outcome interviews of 51 patients was of sufficiently high quality to be observed and rated by clinical personnel. The average ‘live’ interview HAMD score for the sample of 51 patients was 23.5 + 4.5 (± S.D.). Each set of patient video interviews conducted in the study proper was rated on the VIBES (Katz et al., 1989) by two clinical observers, selected randomly from a group of six psychiatric residents trained in the rating procedures. The residents had no prior contact with the patients and were recruited after the completion of the study proper. The quality of the videotaping for each patient (clarity of picture and sound, number of missing interviews, etc.) was evaluated independently of the study by clinical personnel having no knowledge of the study treatments in order to avoid any potential bias in screening. The large majority of exclusions (14/19) were due to inaudibility (>20% of the interview) of the baseline and/or outcome sessions, critical to the assessment procedure. The others were excluded because of missing interim interviews or seriously impaired visual quality or temporary technical problems with the equipment. These problems were attributed to the newness of the effort at the two institutions and to the state-of-the-art technology, problems easily overcome by new technology in future studies.

Baseline period

Study subjects were treated on an in-patient research unit for 6 wk. All patients were maintained on placebo during an additional 1-wk drug ‘wash-out’ period prior to beginning a 6-wk treatment period. Use of a placebo run-in period in this study was primarily carried out to eliminate any pre-existing drug that could confound identification of early behavioural changes related to the study medications. Clinical response at the end of the pre-treatment period was measured using the HAMD and by a ‘multivantaged’ set of behavioural techniques (Katz et al., 2004a). Regardless of the initial level of severity, a patient was not included in the treatment study if by day 7 the patient no longer scored ≥18 on the HAMD.
Treatment period

After the 7-d pre-treatment period, patients were randomly assigned to either DMI, paroxetine, or placebo. The dosage of DMI was started at 50 mg and was raised as necessary to a maximum of 350 mg/d to reach a blood level of \( \geq 125 \text{ ng/ml} \) by day 7. Steady-state concentrations of DMI were reached by day 13 for 80% of the patients in the study. Twenty-nine were assigned to treatment with DMI. Of the initial patient sample entered, three dropped out within 2 wk due to side-effects. Of the remaining 26 patients, 16 were included in the video study. In all 28 patients were assigned to paroxetine. Of these, four did not complete at least 3 wk of the protocol. The dosage for paroxetine ranged from 20 to 60 mg/d. The dosage was adjusted to achieve a minimum steady-state serum concentration of drug of at least 10 ng/ml. Steady-state concentrations of paroxetine were reached by day 6 for 94% of the patients. Of the 28 patients assigned to paroxetine 18 were included in the video study. A total of 25 patients were assigned to placebo in the study proper and five did not complete 3 wk of treatment. Of the 20 remaining patients, 17 were in the video study. Plasma levels of drug were monitored daily during the first week of the study, twice weekly for the next 2 wk, then once a week until the end of the study. Thus, we were able to ensure equivalent treatment in the active medication groups.

As the focus of the study was on early behavioural change, it was decided a priori that patients who did not complete at least 3 wk of treatment would not generate useful data for the analyses. Of the 51 patients included in the data analyses in the video study, however, two completed 5 wk treatment and two completed 3 wk. For such patients their last behavioural scores were considered to be their final score and they were assigned a behavioural outcome based on these scores.

Measuring the components and severity of the depressed state

To address the questions and hypotheses raised in the study proper, established measures of specific behavioural components as well as the severity of the disorder were required (Katz et al., 2004b). In this video study the focus was on the VIBES, the expressive and behavioural components that evolved from prior psychometric analyses and the new severity dimensions that had been developed for its use in clinical trials.

Behavioural components

To adapt the method for application in clinical trials, the VIBES standardized mental status interview and the scales were reduced in length from 35 to 24 questions, reducing the average interview time from 20 to 10 min. The original VIBES expressive and symptom scales were developed to be applicable to a range of disorders, including schizophrenia (Katz and Itil, 1974). The currently applied method was designed, however, to target affective disorders (Katz et al., 1989). The scoring system was based on data from a primarily, depressed population of patients (Katz et al., 1984). It only utilized 53 of the original 88 expressive and symptom items and 33 of the 66 items in the social behaviour scales. In the new Brief form these currently unscored items were dropped, as were five others that contributed minimally to the set of constructs. The Brief version constructs are, therefore, scored, except for a few minor items, in the same manner as the earlier longer version.

The observational and rating procedure for an interview was thereby reduced from \( \sim 30 \text{ min to 12 min} \). Most all the items that comprised the 11 expressive and symptom components of the VIBES and the eight social behaviour factors were retained (Katz et al., 1989).

The VIBES expressive factors include retardation of movement and speech, agitation, distressed facial expression, bodily tension, and detachment-indecisiveness; symptom scales include depressed mood, anxiety, hostility–disturbed judgement, somatization, cognitive impairment, and apathy-confusion. The social behaviour scales include measures of positive social adaptation, anxiety, physical agitation, distractibility, suspiciousness, openness, irritability and verbal aggression.

Frequency of measurement

Behavioural measurements were administered twice weekly during the first 3 wk of treatment and weekly from weeks 4–6. The patients were video interviewed in each assessment period utilizing the standardized mental status question schedule from the VIBES method (Katz et al., 1989). Based on observation of that recorded interview an independent rater (psychiatric resident) completed the Brief expressive and symptom scales of the VIBES and the social behaviour scales during the course of the interview and on a global severity scale at the conclusion of the interview. The scales were calculated with a computer program from the database. The entire interview and rating procedure for an interview took \( \sim 12 \text{ min} \). Each patient...
was assessed at six time-points: baseline (pre-treatment), during treatment (at days 7, 10, 13 and 16) and at outcome (42 d). Observation and rating of the six interviews were conducted in 60–70 min.

Measuring outcome
Outcome was measured at the end of 6 wk of treatment using the same categorical index (Maas et al., 1984) for assigning patients that was used in the Collaborative Study in-patient study. It is based on combined criteria from improvement and final status ratings on the HAMD, CGI (Guy 1976), and Global Assessment Scale (GAS; Endicott et al., 1976). This index divided patients into ‘clear responders’, ‘indeterminates (partial responders)’, and non-responders. Because of the small number of categorical non-responders in the drug groups, the course of drug action was compared with that of the combined indeterminate and non-responders, designated as ‘non-responders’. In this video study, for DMI there were 10 responders, 6 non-responders; for paroxetine, the respective categories were 9 and 8, and for placebo, 6 and 8.

A second outcome based on the HAMD where response is defined as a 50% decrease in the total (21-item) score, was used also.

Statistical methods
The plan of analysis was first to compare HAMD severity results based on ‘live’ interviews in the main study with VIBES global severity and severity dimensions in this study. This step is followed by analysis of drugs vs. placebo actions on the VIBES factors.

Derivation of the VIBES severity dimensions and reliability of factors
To derive the independent dimensions of severity measured by the 11 VIBES scales of expressivity and symptoms and the eight social behaviour scales, a principal components analysis with normalized varimax rotation (Morrison, 1967) was applied to the pre-treatment (baseline) values of the 51 patients. Patients were then scored on the new derived severity dimensions.

The inter-rater reliabilities of the VIBES factors and derived severity dimensions were assessed using the intra-class correlation (ICC) on the ratings by two observers of each patient.

Performance of the video method
Analysis of treatment efficacy
To compare the video method with the HAMD, both the HAMD outcome index of at least a 50% reduction in the total score and the categorical outcome measures, based on the ‘live’ interviews, were used. \( \chi^2 \) and Fisher’s exact tests were used to compare the proportions of responders to DMI and paroxetine at outcome with the proportion of placebo-treated responders. Analysis of covariance (ANCOVA) on the final value of each measure (day 42), with treatment as the independent group, and covarying for the baseline value, was used to assess the absolute difference between drug groups after 6 wk of treatment on the video severity dimension and component measures and on the ‘live’ HAMD values.

Analysis of early behavioural actions and onset comparing treatment results in all subjects
To compare the two methods, analyses of the sensitivity of the VIBES severity dimensions and of the HAMD scale in detecting onset and early treatment-induced changes in drug responders were first conducted separately. To measure onset and rate of action, a mixed model analysis of variance (ANOVA, Slopes test; Laird and Ware, 1982) was applied to determine whether the rate of reduction in severity in one group was significantly more rapid during the first 16 d of treatment than for the other treatment groups. If a difference in the rate of reduction was shown, then for each measure and treatment, ANCOVA at each time-point covarying for the baseline value was run to identify the initial time-points at which the reduction of severity was significantly different between groups. In analysing results using the Brief VIBES, the focus was first on the four severity dimensions, and second on factors found to change within the first 2 wk in previous studies (Katz et al., 1987, 2004b). These factors included depressed mood, anxiety, hostility, motor retardation and distressed expression.

In order to identify early clinical actions that were specifically ‘therapeutic’, time-course and change for treatment responders were compared directly with those of non-responders to that treatment, within treatment groups using \( t \) tests at each time-point. To measure the time-course of change for treatment responders on the Brief VIBES and the severity of specific variables over the entire treatment period survival analysis was also applied. The onset of ‘improvement’ for a particular variable was defined according to Stassen et al. (1993), as the initial
time-point at which a reduction of $\geq 20\%$ occurs, which is then sustained throughout the course of the treatment. Onset of ‘full response’ was defined as the time-point when a sustained reduction of $\geq 50\%$ occurs.

**Results**

**Derivation of VIBES severity dimensions**

Utilizing the values from the baseline (pre-treatment) data on the 51 study patients, the intercorrelations of the 11 VIBES expressive symptom and eight social behaviour factors were computed. A principal components analysis of the VIBES scales identified four orthogonal dimensions, accounting for 68\% of the variance. The dimensions, which consist of averages of factors, represent independent components of the severity of the disorder and help to explain the major areas of functioning measured by the VIBES. The four are: (i) social withdrawal–retardation (combined social behaviour and motor activity factors); (ii) agitation–anxiety; (iii) hostility; (iv) depressed mood–cognitive impairment; the latter dimension apparently reflecting core symptoms of the depressive disorder.

**Inter-rater reliability of the revised Brief factors and the severity dimensions**

Inter-rater reliabilities for the Brief VIBES factors and dimensions were assessed with the ICC in the same manner in which the original VIBES factors had been measured. The new Brief factors, however, were tested on a sample of 51 patients compared to the sample of 102 patients in the original (Katz et al., 1984). It could, therefore, be expected that the reliabilities would not be as high for the brief as for the original constructs. The reliability coefficients for the expressive scales in the Brief method ranged from 0.35 to 0.71, the symptom components, from 0.37 to 0.80 and the social behavioural scales from 0.38 to 0.73 (see Table 1). The ICCs for the severity dimensions were, as expected, generally higher with a range of 0.48 to 0.75. The reliabilities of the Brief VIBES factor and severity dimensional scales are presented in Table 1. Although the reliabilities of the Brief factors are not as high as the original factors, they range from moderate to substantial (Landis and Koch, 1977), and meet adequacy standards for psychiatric rating methods.

**Efficacy of DMI and paroxetine vs. placebo at 6 wk: comparison of the VIBES and the ‘live’ HAMD**

These analyses were based on the 51 patients in the current video study. On the $\geq 50\%$ improvement HAMD rating scale criterion used in the ‘live’ interviews in the main study to assess change at 6 wk, DMI resulted in a higher proportion of recovered patients, 65\% meeting the ‘recovery’ criterion to 38\% for the placebo-treated group. The difference, however, unlike the result from the completed sample of 70 in

<table>
<thead>
<tr>
<th>VIBES components</th>
<th>Expressive scales</th>
<th>Symptom scales</th>
<th>Social behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor retardation</td>
<td>0.57</td>
<td>Depressed mood</td>
<td>Positive adaptation</td>
</tr>
<tr>
<td>Distressed exp.</td>
<td>0.71</td>
<td>Anxiety</td>
<td>Irritability</td>
</tr>
<tr>
<td>Bodily tension</td>
<td>0.65</td>
<td>Somatization</td>
<td>Agitation</td>
</tr>
<tr>
<td>Detached–indecisive</td>
<td>0.35</td>
<td>Cognition</td>
<td>Distraction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hostility</td>
<td>Suspicious</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Apathy–Confusion</td>
<td>Openness</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nervous</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Verbal aggression</td>
</tr>
</tbody>
</table>

*Based on interview and paired ratings of 46 depressed patients.
Early behavioural actions of DMI, paroxetine and placebo

The HAMD was able to detect significantly greater improvement in severity of the disorder at 7 d with DMI ($p < 0.05$) when compared with placebo and paroxetine, as was the VIBES measure of global severity ($p < 0.05$). The Slopes test indicated significantly more rapid DMI-induced reduction of VIBES measures of agitation ($p < 0.001$), cognitive impairment ($p < 0.01$), depressed mood ($p < 0.05$) and hostility ($p < 0.05$), and on the severity dimensions, depressed mood–cognitive impairment and agitation–anxiety during the first 16 d than that due to placebo. These comparisons showed DMI to also act more rapidly than paroxetine on all these factors, except hostility.

At 13 d of treatment, DMI was found to significantly improve cognitive impairment ($p < 0.01$), apathy–confusion ($p < 0.01$), and depressed mood when compared with placebo and paroxetine, and when the factors were combined to form the depressed mood–cognitive impairment core depressive dimensional measure ($p < 0.005$) (Table 2, Figure 1).

The ‘live’ HAMD showed a significantly greater reduction in overall severity with DMI than with placebo or paroxetine at day 13.

No differences in rate of action were found between paroxetine and placebo on the dimensions or constructs.

Therapeutic actions

Relationship of early behavioural changes to recovery status at 6 wk

These analyses compared the early behavioural changes in clear responders to those treated with the same drug who were non-responders at 6 wk. For DMI-treated patients, responders showed significantly lower values at 7 d on the VIBES measure of global severity (and through days 10, 13, and 16), at 10 d, in motor retardation ($p < 0.05$) and by day 13 in bodily tension ($p < 0.01$), than in the non-responder group. The hostility dimension was significantly lower in DMI responders at day 13 ($p < 0.05$).

For paroxetine responders significantly lower values than those shown by non-responders were seen as early as day 7 on global severity (sustained at days 10 and 13) and on detachment-indecisive ($p < 0.05$) and apathy–confusion ($p < 0.05$). Placebo responders showed only one difference from non-responders at day 10 in reduction on somatization ($p < 0.05$).

Results indicate that video ratings were sensitive to change in overall severity in DMI responders as early

The analysis of the VIBES factors also showed DMI to be superior to both paroxetine and placebo in specifically, reducing agitation ($p < 0.05$) and bodily tension ($p < 0.01$) at 6 wk (Table 2, Figure 2).

Thus, despite the relatively small sample of treated patients, the VIBES measures were able to detect the efficacy of DMI in comparison to placebo, in reducing the core depressive dimension, and in its specific actions in reducing agitation and physical tension at 6 wk of treatment.

The analysis of the VIBES severity dimensions showed DMI to reduce both the core depressive mood-cognitive impairment (the most similar of the dimensions to the HAMD total score measure) dimension significantly more than paroxetine and placebo at 6 wk ($p < 0.05$). The effect on depressed mood-cognitive impairment is shown in Figure 1.

The analysis of the VIBES factors also showed DMI to be more rapidly during the first 16 d of treatment than did placebo or paroxetine (Slopes test: ** $p < 0.01$, * $p < 0.05$ respectively). DMI reduced DM–CI significantly more than did placebo or paroxetine at day 13 (ANCOVA, ** $p < 0.01$), day 16 (* $p < 0.05$), and at outcome (* $p < 0.05$) of treatment.

Figure 1. VIBES depressed mood–cognitive impairment (DM–CI): comparing desipramine (DMI; – ■ –), paroxetine (– –) and placebo (– ▲ –) treatments across time-points. DMI reduced severity dimension, DM–CI, significantly more rapidly during the first 16 d of treatment than did placebo or paroxetine (Slopes test: ** $p < 0.01$, * $p < 0.05$ respectively). DMI reduced DM–CI significantly more than did placebo or paroxetine at day 13 (ANCOVA, ** $p < 0.01$), day 16 (* $p < 0.05$), and at outcome (* $p < 0.05$) of treatment.

The main study (62% to 30%) (Katz et al., 2004b), was not statistically significant. The percentage of patients treated with paroxetine who had this level of improvement (46%) was also not significantly different from the proportions for either DMI or placebo. Similarly, when ANCOVA was applied to an analysis of HAMD total scores at 6 wk, covarying for baseline ratings, DMI was not found to have reduced overall severity of depression significantly more than placebo, and no differences were found between paroxetine and placebo or between the two drugs.

Analysis of the VIBES severity dimensions showed DMI to reduce the core depressive mood-cognitive impairment (the most similar of the dimensions to the HAMD total score measure) dimension significantly more than paroxetine and placebo at 6 wk ($p < 0.05$). The effect on depressed mood-cognitive impairment is shown in Figure 1.

The hostility dimension was significantly lower in DMI responders at day 13 ($p < 0.05$). For paroxetine responders significantly lower values than those shown by non-responders were seen as early as day 7 on global severity (sustained at days 10 and 13) and on detachment-indecisive ($p < 0.05$) and apathy–confusion ($p < 0.05$).

Placebo responders showed only one difference from non-responders at day 10 in reduction on somatization ($p < 0.05$).

Results indicate that video ratings were sensitive to change in overall severity in DMI responders as early

the main study (62% to 30%) (Katz et al., 2004b), was not statistically significant. The percentage of patients treated with paroxetine who had this level of improvement (46%) was also not significantly different from the proportions for either DMI or placebo. Similarly, when ANCOVA was applied to an analysis of HAMD total scores at 6 wk, covarying for baseline ratings, DMI was not found to have reduced overall severity of depression significantly more than placebo, and no differences were found between paroxetine and placebo or between the two drugs.

Analysis of the VIBES severity dimensions showed DMI to reduce both the core depressive mood-cognitive impairment (the most similar of the dimensions to the HAMD total score measure) dimension significantly more than paroxetine and placebo at 6 wk ($p < 0.05$). The effect on depressed mood-cognitive impairment is shown in Figure 1.

The analysis of the VIBES factors also showed DMI to be more rapidly during the first 16 d of treatment than did placebo or paroxetine (Slopes test: ** $p < 0.01$, * $p < 0.05$ respectively). DMI reduced DM–CI significantly more than did placebo or paroxetine at day 13 (ANCOVA, ** $p < 0.01$), day 16 (* $p < 0.05$), and at outcome (* $p < 0.05$) of treatment.

Figure 1. VIBES depressed mood–cognitive impairment (DM–CI): comparing desipramine (DMI; – ■ –), paroxetine (– –) and placebo (– ▲ –) treatments across time-points. DMI reduced severity dimension, DM–CI, significantly more rapidly during the first 16 d of treatment than did placebo or paroxetine (Slopes test: ** $p < 0.01$, * $p < 0.05$ respectively). DMI reduced DM–CI significantly more than did placebo or paroxetine at day 13 (ANCOVA, ** $p < 0.01$), day 16 (* $p < 0.05$), and at outcome (* $p < 0.05$) of treatment.

The main study (62% to 30%) (Katz et al., 2004b), was not statistically significant. The percentage of patients treated with paroxetine who had this level of improvement (46%) was also not significantly different from the proportions for either DMI or placebo. Similarly, when ANCOVA was applied to an analysis of HAMD total scores at 6 wk, covarying for baseline ratings, DMI was not found to have reduced overall severity of depression significantly more than placebo, and no differences were found between paroxetine and placebo or between the two drugs.

Analysis of the VIBES severity dimensions showed DMI to reduce both the core depressive mood-cognitive impairment (the most similar of the dimensions to the HAMD total score measure) dimension significantly more than paroxetine and placebo at 6 wk ($p < 0.05$). The effect on depressed mood-cognitive impairment is shown in Figure 1.

The analysis of the VIBES factors also showed DMI to be superior to both paroxetine and placebo in specifically, reducing agitation ($p < 0.05$) and bodily tension ($p < 0.01$) at 6 wk (Table 2, Figure 2).

Thus, despite the relatively small sample of treated patients, the VIBES measures were able to detect the efficacy of DMI in comparison to placebo, in reducing the core depressive dimension, and in its specific actions in reducing agitation and physical tension at 6 wk of treatment.

Early behavioural actions of DMI, paroxetine and placebo

The HAMD was able to detect significantly greater improvement in severity of the disorder at 7 d with DMI ($p < 0.05$) when compared with placebo and paroxetine, as was the VIBES measure of global severity ($p < 0.05$). The Slopes test indicated significantly more rapid DMI-induced reduction of VIBES measures of agitation ($p < 0.001$), cognitive impairment ($p < 0.01$), depressed mood ($p < 0.05$) and hostility ($p < 0.05$), and on the severity dimensions, depressed mood–cognitive impairment and agitation–anxiety during the first 16 d than that due to placebo. These comparisons showed DMI to also act more rapidly than paroxetine on all these factors, except hostility.

At 13 d of treatment, DMI was found to significantly improve cognitive impairment ($p < 0.01$), apathy–confusion ($p < 0.01$), and depressed mood when compared with placebo and paroxetine, and when the factors were combined to form the depressed mood–cognitive impairment core depressive dimensional measure ($p < 0.005$) (Table 2, Figure 1).

The ‘live’ HAMD showed a significantly greater reduction in overall severity with DMI than with placebo or paroxetine at day 13.

No differences in rate of action were found between paroxetine and placebo on the dimensions or constructs.

Therapeutic actions

Relationship of early behavioural changes to recovery status at 6 wk

These analyses compared the early behavioural changes in clear responders to those treated with the same drug who were non-responders at 6 wk. For DMI-treated patients, responders showed significantly lower values at 7 d on the VIBES measure of global severity (and through days 10, 13, and 16), at 10 d, in motor retardation ($p < 0.05$) and by day 13 in bodily tension ($p < 0.01$), than in the non-responder group. The hostility dimension was significantly lower in DMI responders at day 13 ($p < 0.05$).

For paroxetine responders significantly lower values than those shown by non-responders were seen as early as day 7 on global severity (sustained at days 10 and 13) and on detachment-indecisive ($p < 0.05$) and apathy–confusion ($p < 0.05$). Placebo responders showed only one difference from non-responders at day 10 in reduction on somatization ($p < 0.05$).

Results indicate that video ratings were sensitive to change in overall severity in DMI responders as early
as 7 d of treatment, that specific behavioural changes were detected in the reduction of motor retardation at 10 d, and the relaxation of agitation and bodily tension at 2 wk, when compared with patients who were non-responsive to the drug.

For paroxetine, early changes in clearing confused thinking and indecisiveness may relate indirectly to a reduction in anxiety. The time-course curves for the clinical actions in treatment responders to DMI, paroxetine and placebo were examined with survival

Table 2. DMI vs. paroxetine vs. placebo: comparison of changes at 1, 2 and 6 wk of treatment

<table>
<thead>
<tr>
<th>VIBES factors</th>
<th>Week 0</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 6</th>
<th>p value</th>
<th>ANOVA results* Differences in change</th>
<th>Slopes test results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DMI vs. paroxetine</td>
<td>DMI vs. placebo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p value</td>
<td>p value</td>
</tr>
<tr>
<td>Depressed mood</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMI</td>
<td>2.55 (0.15)</td>
<td>2.10 (0.16)</td>
<td>1.80 (0.14)</td>
<td>1.62 (0.13)</td>
<td>0.0330</td>
<td></td>
<td>0.0129</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>2.73 (0.11)</td>
<td>2.36 (0.09)</td>
<td>2.26 (0.13)</td>
<td>1.92 (0.14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>2.73 (0.13)</td>
<td>2.42 (0.13)</td>
<td>2.29 (0.13)</td>
<td>1.99 (0.13)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMI</td>
<td>2.44 (0.19)</td>
<td>2.08 (0.23)</td>
<td>1.82 (0.16)</td>
<td>1.54 (0.13)</td>
<td></td>
<td></td>
<td>0.0112</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>2.44 (0.13)</td>
<td>2.32 (0.1)</td>
<td>1.95 (0.1)</td>
<td>1.87 (0.12)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>2.38 (0.15)</td>
<td>2.26 (0.15)</td>
<td>2.12 (0.13)</td>
<td>1.87 (0.14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMI</td>
<td>1.71 (0.12)</td>
<td>1.45 (0.11)</td>
<td>1.34 (0.08)</td>
<td>1.14 (0.06)</td>
<td>0.0330</td>
<td></td>
<td>0.0001</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>1.519 (0.09)</td>
<td>1.417 (0.07)</td>
<td>1.403 (0.05)</td>
<td>1.298 (0.06)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>1.368 (0.08)</td>
<td>1.333 (0.07)</td>
<td>1.294 (0.07)</td>
<td>1.258 (0.05)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bodily tension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMI</td>
<td>1.79 (0.1)</td>
<td>1.54 (0.09)</td>
<td>1.52 (0.11)</td>
<td>1.26 (0.07)</td>
<td>0.0090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paroxetine</td>
<td>1.98 (0.11)</td>
<td>1.83 (0.1)</td>
<td>1.68 (0.08)</td>
<td>1.63 (0.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>1.85 (0.09)</td>
<td>1.79 (0.08)</td>
<td>1.61 (0.08)</td>
<td>1.50 (0.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive impairment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMI</td>
<td>2.27 (0.16)</td>
<td>1.95 (0.16)</td>
<td>1.52 (0.11)</td>
<td>1.39 (0.1)</td>
<td>0.0015</td>
<td></td>
<td>0.0335^a</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>2.52 (0.13)</td>
<td>2.39 (0.12)</td>
<td>2.15 (0.11)</td>
<td>1.89 (0.13)</td>
<td></td>
<td></td>
<td>0.0102</td>
</tr>
<tr>
<td>Placebo</td>
<td>2.45 (0.12)</td>
<td>2.32 (0.11)</td>
<td>2.07 (0.14)</td>
<td>1.78 (0.12)</td>
<td></td>
<td></td>
<td>0.0101</td>
</tr>
<tr>
<td>Apathy–Confusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMI</td>
<td>1.57 (0.1)</td>
<td>1.39 (0.07)</td>
<td>1.25 (0.08)</td>
<td>1.19 (0.06)</td>
<td>0.0270</td>
<td></td>
<td>0.0085</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>1.65 (0.1)</td>
<td>1.58 (0.08)</td>
<td>1.53 (0.08)</td>
<td>1.36 (0.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>1.80 (0.1)</td>
<td>1.71 (0.1)</td>
<td>1.57 (0.09)</td>
<td>1.46 (0.09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIBES Severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMI</td>
<td>4.00 (0.75)</td>
<td>3.50 (0.89)</td>
<td>3.06 (0.93)</td>
<td>2.50 (0.96)</td>
<td>0.0338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paroxetine</td>
<td>4.56 (0.95)</td>
<td>4.28 (0.81)</td>
<td>3.89 (0.96)</td>
<td>3.25 (1.18)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>4.47 (0.67)</td>
<td>4.27 (0.71)</td>
<td>3.88 (0.7)</td>
<td>3.32 (0.92)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAMD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMI</td>
<td>16.00 (1.43)</td>
<td>24.06 (1.59)</td>
<td>18.94 (1.64)</td>
<td>15.31 (1.49)</td>
<td>0.0281b</td>
<td></td>
<td>0.0304b</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>24.00 (1.14)</td>
<td>21.17 (1.08)</td>
<td>20.17 (1.12)</td>
<td>18.44 (1.26)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>25.47 (0.99)</td>
<td>20.76 (1.16)</td>
<td>19.94 (1.18)</td>
<td>16.00 (1.31)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VIBES, Video Interview Behaviour Evaluation Scales, DMI, desipramine; HAMD, Hamilton Depression Rating Scale.

*Pairwise comparisons show DMI < paroxetine and placebo except where noted.

^a DMI < Paroxetine only.

b DMI < Placebo only.

as 7 d of treatment, that specific behavioural changes were detected in the reduction of motor retardation at 10 d, and the relaxation of agitation and bodily tension at 2 wk, when compared with patients who were non-responsive to the drug.

For paroxetine, early changes in clearing confused thinking and indecisiveness may relate indirectly to a reduction in anxiety. The time-course curves for the clinical actions in treatment responders to DMI, paroxetine and placebo were examined with survival
analysis. Despite the relatively small sample sizes, significant differences were found for the rates of reduction on agitation \((p < 0.01)\), bodily tension \((p < 0.01)\), detachment \((p < 0.05)\), and anxiety \((p < 0.05)\). All differences showed DMI reductions to be more rapid than those induced by paroxetine or placebo. These results do not appear to differ meaningfully from those obtained with the ANCOVA analyses.

**Discussion**

It is clear that the development of new types of ADs, more broadly effective and more rapidly acting than the current ones, has seriously lagged since the introduction of the SSRIs in the early 1980s. This is due in part to our incomplete understanding of how drug-induced changes in neurotransmitter systems translate into specific changes in behaviour, mood and cognition. One of the limitations to finding new drugs is the failure to use refined behavioural methods which assess not only changes in severity of the disorder, but uncover the initial actions on facets of the disorder, e.g., on motor activity, mood and cognitive functioning. We have shown in recent studies how such methods can enhance understanding of the sequence of behavioural changes that underlie therapeutic efficacy, i.e., the manner in which drugs bring about an ‘anti-depressant’ result (Katz et al., 2004a,b). Such findings also lead directly to identifying potential applications for the drug’s use in treating other disorders.

**The VIBES as an evaluation method in clinical studies**

Assessment of the effectiveness of a new, putative AD in reducing the severity of a depressive disorder still relies primarily on the judgements of a trained clinical observer. The evaluation is based on the application of established symptom rating scales, such as the HAMD or the MADRS (Montgomery and Asberg, 1979), to ‘live’ mental status interviews with the patient, conducted before treatment, during and at outcome. Video recording of such interviews has certain obvious advantages over simply relying on the ‘live’ interview. It permits increased reliability of ratings through use of multiple observers, greater detachment on the part of the observer, a focus on physical and social behaviour, and the opportunity to juxtapose baseline and outcome interviews thereby reducing the role of memory in the observer’s judgement. For experimental purposes, it makes possible a more refined analysis of the actual timing and sequence of drug-induced behavioural changes. Other practical advantages are the provision of a permanent record of the patient’s progress, and the added convenience of greater flexibility in the scheduling of assessments for clinical personnel.

Many years of experience in trying to adapt video methodology to clinical trials, however, turned up significant problems that have restricted its use to training situations and prevented its application to controlled research. A significant problem has been the cost and the awkwardness of the technology. The prime drawbacks, however, have been the length of the interviews, the tedium of observing and rating 30- to 40-min interviews and the lack of behavioural methods designed to take advantage of the increased objectivity that observer detachment makes to the rating situation. Video technology has since improved markedly and the need for technical assistance has been virtually eliminated. In view of the limited applicability of the original VIBES (Katz et al., 1989), the standard mental status interview has been markedly shortened (now 12–15 min) and the expressive, symptom and social behaviour rating scales reduced in length by ~50%. It is now possible to view and rate a patient’s before, during-treatment and outcome interviews in ~1 h. It was then feasible to test the revised VIBES method in a clinical study which compared an established selective noradrenergic and a serotonergic AD with placebo, using established measures of severity, e.g., the HAMD, and of behaviour, to evaluate efficacy and early behavioural changes (Katz et al., 2004b). The main study had sufficient
power to show differences between drugs and placebo. DMI was found, through assessment of ‘live’ interviews, to be clearly more effective than placebo and the drugs to result in different behavioural changes within the first 2 wk of treatment.

The aims of the current study were to determine if the VIBES method was as sensitive in assessing efficacy and in detecting onset and early behavioural actions of ADs, as the standard procedure in the field, i.e. use of the HAMD in before and after ‘live’ interviews.

Regarding efficacy, although the ‘live’ HAMD in the main study showed DMI to be superior to placebo at outcome, the difference between the drug and placebo on the HAMD total scores at outcome, was, unlike the VIBES finding, not significant in the smaller sample video study. On the other hand, the HAMD was as sensitive in the current study to changes in early behavioural actions of the two drugs as it was in the main study. The findings led to the following conclusions:

- The VIBES global severity scale, in a relatively small sample clinical trial, was more sensitive than the HAMD in assessing overall efficacy of an established AD, DMI, when compared with placebo.
- In addition to providing a measure of change in overall severity of the disorder, the VIBES expressive, social behavioural, and mood measures permit refined analysis of the clinical actions of potential AD agents. In this study, DMI was shown to reduce agitation and bodily tension at treatment outcome significantly more than placebo or paroxetine.
- On the detection of onset and early drug actions, both the ‘live’ HAMD and the VIBES were shown to be sufficiently sensitive to detect onset of DMI actions through change in overall severity as early as 7 d, when compared with placebo and paroxetine.
- In further elaborating the earlier drug actions, the VIBES was able to detect more rapid action by DMI in reducing cognitive impairment, apathy and confusion, and depressed mood, during the first 2 wk, than did placebo and paroxetine.

The 2-wk changes for DMI, i.e. the elevation of mood and increased cognitive clarity, appeared to reflect an initial ‘stimulating’ action of the drug, preceding the reductions detected at day 16 of agitation and bodily tension. The later actions suggest a secondary ‘calming’ or ‘tranquilizing’ effect of the drug. Although the time between days 13 and 16 is short, the sequence of behavioural changes is the same as that found in the main study where a more established set of behavioural methods was used.

- Regarding ‘therapeutic actions’, the time-course of clinical actions in ‘responders’ were compared with that of patients who show little or no response. The initial specific actions in the responders, were in reducing motor retardation at 10 d, and then at 2 wk, in reducing agitation and bodily tension. The sequence of stimulatory and calming effects were consonant with those found with DMI in responders in the main study which used more elaborate behavioural methodology to show a strong initial action on motor activity (Katz et al., 2004b). The effects in paroxetine responders were greater by day 7 on overall severity, on indecisiveness, and on apathy–confusion. These effects were not the same as those found in the main study in which an initial reduction in anxiety was shown by day 10. In view of the association of indecisiveness to anxiety, however, the effects found in the two analyses could have a similar meaning.

VIBES as a method for assessing efficacy of an established AD in a relatively small sample study, is more sensitive than the HAMD. Further, its capacity to provide measures of the specifics of behavioural change, particularly motoric and social behavioural aspects, uncovers other important changes at outcome that can be traced to the drug’s action. Thus, the profile of behavioural change that is associated with the drug’s clinical actions can be specified, indicating that the drug may be applied to the treatment of other disorders, e.g. generalized anxiety disorder. Regarding detection of onset and early behavioural actions, the VIBES is as sensitive as the HAMD in assessing the rate of AD action. That is shown in the method’s capacity to detect significant impact on the disorder as early as 7 d. The findings go further than the HAMD, however, in showing DMI to reduce certain symptomatic behaviours at a more rapid rate than produced in placebo-treated patients.

**Application in multicentre or multinational trials**

Although apparently effective for application in clinical trials, the practicality of the VIBES for multicentre drug trials depends mainly on the staff time required for the evaluation process. The VIBES standard interview is similar in length to the semi-structured HAMD interview but the ratings take slightly longer to administer. The evaluation time, therefore, depends mainly on whether the study requires the pre-drug and outcome interviews only, or is also designed to assess change during treatment, requiring interim interviews. Regarding efficiency of evaluation and cost
saving, once the tapes are collected, all can be viewed by clinical staff at a central location. Use of raters already trained to conduct highly reliable assessments would eliminate the variance attributable to merging ratings across centres. In addition, providing a live record of all evaluation data makes possible the conduct of follow-up studies, e.g. of the responsiveness of specific subtypes of depression or of the early behavioural changes shown by treatment-responsive patients. Thus, the video library provides opportunities in clinical research, not available with other methodologies. If applied in international studies, the method requires translation and then demonstration that it is valid in non-English-speaking settings. Such applications have been conducted with the HAMD and other similar methods. In view of the similarity of the VIBES in content to these other approaches, there is good reason to believe translated forms of the VIBES would be equally effective.

In conclusion, the VIBES, a technology for assessing severity and for characterizing the behavioural psychopathology of depressed patients, is, therefore, capable of providing new and sensitive information on the onset and nature of clinical actions of ADs. The method can enhance the understanding of the sequence of behavioural changes that underlies therapeutic efficacy, i.e. the manner in which drugs bring about an ‘antidepressant’ result. It can lead directly to identifying potential applications for the drug’s use in treating other disorders. We, therefore, recommend wider use of this new behavioural technology in the attempts to develop new drugs for the affective disorders.

Acknowledgements

We thank Valerie Dullnig for monitoring all aspects of the study in San Antonio and Virginia Doyle, Pharm.D., who provided excellent assistance with the study medications. We also acknowledge the study coordinators in Dallas, Greer Garner, Ph.D., Janice Malester, Ph.D., and Evelyn Gaspard-Parker, Ph.D. The work was supported by grants from the Department of Veteran Affairs Merit Award (A.F.) and Friends for Psychiatric Research, Department of Psychiatry, University of Texas Health Science Center at San Antonio (M.M.K.).

Statement of Interest

None.

References


