

How long does it take for antidepressant therapies to act?

Gordon Parker, Kay Roy, David B. Menkes, John Snowdon, Philip Boyce, David Grounds, Bernard Hughson, Catherine Stringer

Objective: To review the proposition that antidepressants have a delayed onset of action by employing measurement and analytic strategies that overcome problems confounding interpretation of many efficacy studies.

Method: A subset of patients was recruited to the longitudinal component of the Australasian database study, was assessed at baseline, and then completed measures of depression and anxiety when treatment commenced, and every 3 days over the next 4 weeks. The trajectories of defined 4-week outcome responders and non-responders were compared.

Results: Both groups showed a similar decrease in depression (and anxiety) over the first 3 days. A clear trend break then occurred, with little further improvement in the non-responders, as against distinct and progressive improvement in the responders. Ongoing early improvement (across days 3–6) was a strong predictor of responder status.

Conclusions: The small sample size limits firm interpretation, although distinct interpretive advantages to the study design are evident. Findings are compatible with a number of recent studies arguing against any extensive delayed onset of action for the antidepressant drugs, but argue for caution in interpreting immediate improvement as predicting likely responder status, and more for examining early and sustained improvement as such a marker.

Key words: antidepressant treatments, depression, treatment effectiveness.

Australian and New Zealand Journal of Psychiatry 2000; 34:65–70

Gordon Parker, Professor of Psychiatry (Correspondence); Kay Roy, Research Assistant

School of Psychiatry, University of New South Wales, Psychiatry Unit, Prince of Wales Hospital, Randwick, New South Wales 2031, Australia. Email: g.parker@unsw.edu.au

David B. Menkes, Liaison Psychiatrist

Dunedin Hospital, Dunedin, New Zealand

John Snowdon, Associate Professor; Philip Boyce, Professor

Department of Psychological Medicine, University of Sydney, Sydney, Australia

David Grounds, Psychiatrist

Melbourne Clinic, Melbourne, Australia

Bernard Hughson, Psychiatrist

Canberra Hospital, Canberra, Australia

Catherine Stringer

North Hobart, Australia

Received 28 May 1999; accepted 18 August 1999.

It is commonly suggested that antidepressant medications have a significantly delayed onset of action. Gershon [1] stated that 'It has been an established belief that pharmacotherapy with antidepressant agents takes several weeks, 2, 4, or 6 to produce their therapeutic activity'. It is not uncommon for pharmaceutical companies to suggest much longer intervals, and clinicians often wait, and encourage their patient to wait, for weeks or months before concluding that an antidepressant drug is ineffective. This generally assumed delayed onset is, in fact, contrary to the initial study describing the benefits of imipramine [2]. The assumption of delayed onset of action has been more recently challenged [3] and as that paper, and other relevant studies, have been reviewed in some detail [4–6] only a summary will be considered here.

Firstly, the issue can become a semantic one. How long it might take an antidepressant therapy to 'act' could depend on whether time to initial improvement, substantial improvement, remission or recovery of episode is being considered. Onset of action is the issue examined here, arguing for evidence of some 'improvement'. Some precision has been brought to the topic by accepting operational definitions. Thus, 'improvement' has been defined [3] as a 20% reduction in depression severity (from initial assessment) at any subsequent stage, and 'response' as a 50% improvement in depression severity by the end of a fixed-interval trial or observation period. However, 'pseudo-improvement' can occur if composite measures of depression are used. For instance, a sedative antidepressant might have some relatively immediate effect on any sleep disturbance, and while it might be sufficient to achieve formal 'improvement' status, improvement may then more reflect a drug side effect rather than any primary effect on mood or hedonic tone. Again, any improvement may not be sustained.

Formalised drug trial efficacy data have perhaps contributed to the impression of a delayed onset of action of antidepressant therapies. Such studies have generally initially compared a new antidepressant with a placebo and demonstrated an improvement trajectory in both those receiving the antidepressant and those receiving the placebo, but as significant differences do not generally appear for several weeks at least, this has caused some to interpret improvement as having been delayed. Drug trials generally measure progress at weekly intervals only, which may have also prevented any evidence of true early improvement from being identified.

Stassen *et al.* [3] argued that the general drug study paradigm of examining grouped data (e.g. total drug A versus total drug B subjects) should be modified, as even clearly grouped data will be made up of disparate groups (including patients who have a rapid remission, those who fail to show any change across the study, and those whose condition worsens). Thus, they argued that such data sets should be deconstructed into separate groups of those who at the end of the trial meet 'responder' or 'non-responder' status. In a study involving three groups (two receiving differing antidepressants and one a placebo), evidence of 'improvement' in responders emerged within the first 5 days, whether subjects were receiving an antidepressant drug or a placebo. This did not necessarily mean that the antidepressant drugs were ineffective. Patients receiving an antidepressant were less likely to 'drop out' of the study and more likely to be in the

'responder' category by the end of the study. This allowed those researchers to suggest that patients with depression may have a 'biological predisposition' to be either a 'responder' or a 'non-responder', and that antidepressant therapy 'essentially converts a percentage of non-responders to responders, so triggering and maintaining the conditions necessary for improvement'.

The possibility that improvement may be evident far earlier than generally claimed has been established for modalities other than antidepressant drugs. Rodger *et al.* [7] established that improvement after the first three treatments of bilateral electroconvulsive therapy (ECT) was six-fold greater than that occurring over the remainder of the course, while Scott and Whalley [8] observed that preliminary studies of ECT neither supported the view that onset of effect is delayed, nor the view that there is little early improvement after ECT initiation.

A Sydney study similarly described early improvement in patients with depression not treated with any antidepressant medication. In a sample of 43 patients assessed initially after referral to a psychiatrist [9], improvement in Zung depression severity scores both at 6 and at 20 weeks was accurately predicted as early as the first assessment which occurred on the sixth day. Additionally, in a sample of untreated subjects with significant depression who volunteered for a research interview but not to receive treatment [10], improvement in Zung depression scores at 6 and at 20 weeks was again predicted by the degree to which the patient had improved by the sixth day. Such studies suggested that those individuals who are going to 'respond' can be expected to show evidence of improvement within the first week.

Such an interpretation does not have universal support. Quitkin, for instance, has long and comprehensively argued for a differing view. In essence, he argued for two broad patterns [11]. First, non-specific improvement, which is associated with early, abrupt or non-persistent initial responses. Second, a true drug response, which is marked by delayed but persistent response. The second interpretation was not necessarily supported by his own data. In an extremely important paper [12], he examined a very large data set involving all studies undertaken over a 10-year period at the New York State Psychiatric Institute. Each contributing study had involved a 7–10 day placebo run-in phase to detect and exclude placebo responders, with progress assessed by weekly ratings. Those who did not show a spontaneous improvement pattern during the run-in phase were

studied weekly, with 6-week double-blind comparisons involving differing antidepressant drugs and the placebo control. Quitkin provided an incidence plot of a 'persistent response' that he interpreted as indicating that 'most of the effect of drug therapies is observed from weeks 3–6'. However, a statistically significant difference in the trajectories is evident at week 2. The key issue in examining databases of this nature is not just to determine when a significant difference is established between the drug and placebo, but more to examine when there is a 'trend break' in the severity estimates of those receiving the active treatment. In his own incidence plot, the trend break was present at the second assessment period at week 2. Thus, his own data showed evidence of early improvement in antidepressant responders.

In the present paper we examine the issue of early improvement in another group heterogeneous in composition and type of antidepressant treatment, derived from the Australasian Data Base (ADB) study. The ADB study design has been described [13]. In essence, it involved 27 Australian and New Zealand psychiatrists contributing 369 patients with depression to a cross-sectional study, wherein the patients and the assessing psychiatrists completed a set of questionnaires with pre-coded rating options, principally assessing current clinical features. In addition, a longitudinal component was provided as an option. While the sub-sample of those followed longitudinally was small, a key advantage was that ratings were made every 3 days and therefore the subsample differs from that used in most studies examining for evidence of early improvement, and where weekly ratings tend to be the strategy. In the light of the small sample size, results are unlikely to be definitive but should serve as a model for considering general issues and methodological nuances important for both clinical interpretation and formal research studies.

Methods

Psychiatrists contributing to the cross-sectional ADB database were also encouraged to have patients enrol in an optional longitudinal component, wherein the patient's progress would be followed over the 4 subsequent weeks with serial and frequent (i.e. third daily) mood ratings. Both at baseline and at follow up, the psychiatrist was required to complete a 17-item Hamilton scale [14], and at follow up complete a seven-item clinical global improvement (CGI) measure rating whether they judged the patient to be worse (assessed across three levels), unchanged or

improved (again assessed across three levels). Those patients who showed a 50% reduction in Hamilton scores across the 4-week interval were regarded as 'responders', with the remaining subjects labelled 'non-responders'. At the final follow up, details were recorded on all interval treatments.

Patients were requested to complete a self-report containing two measures. First, an AUSSI (affect underpinned by severity and social impairment) self-report measure [15] at baseline, every third day and, finally, at follow up. Second, on each occasion subjects were also requested to complete an anxiety measure at the bottom of the AUSSI measure. Here, a single item requested rating 'for today, how anxious (e.g. tense, worried, keyed up and on edge, etc.) you felt at your worst', with anchors of 0 for 'not at all', 1–3 for 'mildly', 4–6 for 'moderately', 7–9 for 'markedly' and 10 for 'extremely'.

Specific treatment details (e.g. dosages of drugs, number of ECT) were not sought.

Results

Thirty-two subjects (62% female; mean age 51.5, SD = 18.9 years) were enrolled in the longitudinal study with 25 returning complete data sets. The remaining subjects had complete data sets up to assessment 7, 8, 9 or 10 (most commonly, it was the final self-report form that was not completed) and, for those patients, their last score was regarded as their follow-up score (i.e. 'last observation carried forward' or LOCF). Thus, all 32 subjects are included in the analyses.

Hamilton and AUSSI scores correlated 0.53 at baseline and 0.69 at follow up respectively ($p < 0.001$). Eighteen (56%) subjects who met the Hamilton definitional requirement of being study 'responders' and all were judged by their psychiatrists as improved on the CGI (i.e. two minimally, eight much and eight very much improved). Fourteen were non-responders and of these nine were judged as improved (one much, eight minimally improved), three were judged unchanged and two were worse.

Mean improvement in CGI scores was 51% greater in the responders than non-responders (5.3 vs 3.5, $t = 6.1$, $p < 0.001$). In the responder group, principal treatments for the 18 were selective serotonin reuptake inhibitor (SSRI) ($n = 8$), ECT ($n = 5$), an antipsychotic ($n = 1$), combination tricyclic/antipsychotic ($n = 1$), a monoamine oxidase inhibitor (MAOI) ($n = 1$) and psychotherapy ($n = 1$), with treatment type not provided for one. In the 14

non-responders, principal treatments were an SSRI ($n = 7$), ECT ($n = 1$), a tricyclic ($n = 1$), an MAOI ($n = 1$), mianserin ($n = 1$) and psychotherapy ($n = 1$), with treatment details not provided for two subjects.

Mean Hamilton scores for the whole sample changed significantly ($t = 6.9$, $p < 0.001$) from baseline (24.3, $SD = 7.7$) to follow up (11.7, $SD = 7.7$), as did mean AUSSI scores (i.e. 22.2, $SD = 7.8$ to 12.1, $SD = 9.7$; $t = 5.6$, $p < 0.001$), respective reductions of 52% and 45%. As shown in Figure 1, the improvement trajectory for the whole group reveals reasonably rapid reduction in AUSSI scores initially, then a small additional improvement between assessments 5 and 7.

Figure 1 also plots the separate trajectories of responders and non-responders, with reduction in their AUSSI depression scores being 67% and 16% respectively, and with final mean Hamilton scores being 6.1 for responders and 18.9 for non-responders.

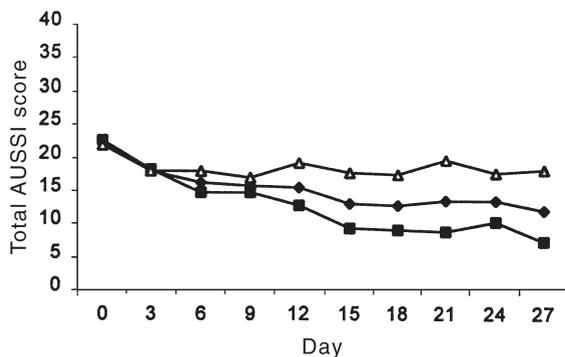


Figure 1. AUSSI total scores from baseline to final assessment: ◆, whole sample; ■, responders; △, non-responders.

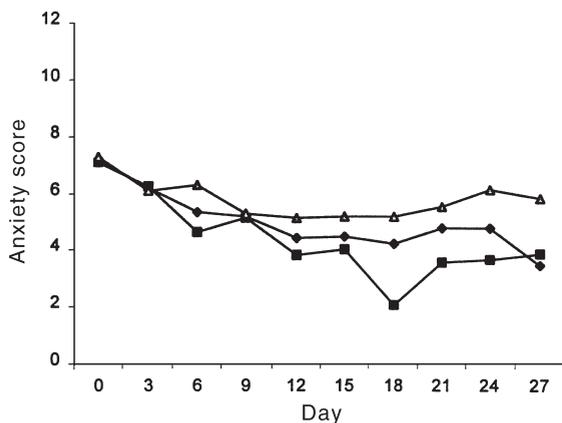


Figure 2. Anxiety scores from baseline to final assessment: ◆, whole sample; ■, responders; △, non-responders.

Those two groups show an identical trajectory from baseline to day 3, but divergence is evident at day 6. The trend break is quite distinctive, with the responders and non-responders establishing quite divergent trajectories by this time, and which were essentially maintained across the rest of the review period.

In Figure 2, third daily anxiety scores are plotted for the whole group and the component responders and non-responders. For the whole group, anxiety levels decreased significantly ($t = 4.3$, $p < 0.001$) by 51%, from a mean of 6.9 ($SD = 2.6$) to 3.4 ($SD = 2.6$). The sub-group patterns are similar to those for the depression data, with both responders and non-responders reporting a similar drop in anxiety in the first 3 days, and then subsequently diverging.

Discussion

As noted, the sample size is small, treatment was not standardised, and differential treatment prevalences across responders and non-responders may have influenced results. The study is then best viewed as a pilot study capable of extension in clinical panel or in formal controlled studies. More importantly, it suggests a paradigm for undertaking studies examining for latency of improvement, and it may assist consideration as to how long patients should receive an antidepressant before considering that it may be ineffective.

The paradigm argues for baseline and regular brief-interval assessment of depression severity, and supports the latter being undertaken by patient self-report. The AUSSI appears acceptable for that purpose, and extends its uses beyond its original objectives [15]. Reduction in AUSSI self-report scores was comparable with reduction in clinician-rated Hamilton scores and those two measures correlated significantly at baseline and at follow up. The advantage of the AUSSI, however, is that it has few items and requires little completion time. Both measures were supported by differential outcomes across responders and non-responders as clinically validated by CGI scores. For research purposes, the suggestion that 'improvement' should not be examined in any overall group, but examined in separate responders and non-responders, was strongly supported. Thus, any overall 'group' change in severity (as reported in most treatment-efficacy studies) is relatively meaningless, being largely influenced by the varying proportions of eventual responders and non-responders.

Despite the small sample size, interesting trends were evident. In both responders and non-responders

a drop in depression and anxiety severity scores appeared of similar magnitude in the first 3 days, raising questions as whether any reduction in anxiety lowered depression or the converse, or a more composite effect. The drop in scores could have been caused by a placebo response and/or a therapeutic effect of baseline assessment itself, or by the fact that the antidepressant interventions were otherwise transiently effective in the subsequent non-responders.

Such a transient global response is important to consider. In the introduction we noted studies challenging a slow onset of action to antidepressant treatments. This could be interpreted as suggesting that early onset of improvement indicates a good prognosis. Certainly, there are a number of studies that argue that proposition. For instance, Priebe and Gruyters [16] analysed data from an inpatient study to argue that improvement in the first 3 days was predictive of outcomes at the time of hospital discharge. Our data clearly show that early improvement (in the first 3 days) occurred as an overall group phenomenon, and therefore, while early improvement would predict 4-week status, any true predictive strength would be weakened by the subsequent lack of ongoing improvement in the non-responders. Such an explanation may hold for many of the studies noted in the introduction, including the Sydney studies of those not receiving any medication [9] and those who merely received a research interview [10].

Progress from days 3–6 was more useful in predicting outcome status, as future non-recovered subjects failed to improve much further from their day 3 status, while responders continued to improve. Quitkin *et al.* [11] held that placebo responders are more likely to show abrupt, early and non-persistent responses. It may be that only the last component has predictive value, for our responders also showed immediate and distinctive improvement. Thus, any generalisation that eventual responders will show early or, conversely, delayed improvement is likely to be flawed. Initial improvement may reflect a definitive antidepressant treatment and/or a non-specific therapeutic ingredient, and is therefore unlikely to be a useful predictor of outcome. Ongoing early improvement (here specifically from days 3 to 6) appears a more substantive predictor of responder status.

Such generalisations from a small database clearly presents risks. There will clearly be some true responders who show a delayed improvement (which may reflect depressive subtype or time required to reach an optimal drug treatment dose or effect), and placebo responders who show evidence of early and sustained

improvement. The extent to which improvement, both as defined here and as differentiating across the responders and non-responders, is clinically meaningful remains somewhat problematic in the same way that drug trial efficacy data do not always correspond with ‘real world’ clinical effectiveness patterns. Nevertheless, our data both challenge and extend the view that antidepressant treatments take some significant time before any improvement can be expected. If the findings are valid, their clinical implications include the possibility of having an estimate of likely responder status relatively early after initiating antidepressant treatment.

Acknowledgments

Funding support for the ADB study was provided by the National Health and Medical Research Council (Program Grant 993208), a New South Wales Health Department Infrastructure Grant Program, and three pharmaceutical companies (Pfizer, Wyeth and Eli Lilly), and study assistance by Heather Brotchie, Kerrie Eyers, Yvonne Foy, Dusan Hadzi-Pavlovic and Christine Taylor.

References

1. Gershon S. Antidepressants: can we determine how quickly they work? *Psychopharmacology Bulletin* 1995; 31:21–22.
2. Kuhn R. Über die behandlung depressiver zustände mit einem Iminopidbenzylderivat. *Schweizerische Medizinische Wochenschrift* 1957; 87:1135–1140.
3. Stassen HH, Delini-Stula A, Angst J. Time course of improvement under antidepressant treatment: a survival-analytical approach. *European Neuropsychopharmacology* 1993; 3:127–135.
4. Parker G. On brightening up: triggers and trajectories to recovery from depression [editorial]. *British Journal of Psychiatry* 1996; 168:263–264.
5. Parker G. Mindset revisited. Recovery from depression: triggers and time patterns. *Australian and New Zealand Journal of Psychiatry* 1996; 30:442–444.
6. Parker G. On lightening up. Improvement trajectories in recovery from depression. *Advances in Psychiatric Treatment* 1996; 2:186–193.
7. Rodger CR, Scott AI, Whalley LJ. Is there a delay in the onset of the antidepressant effect of electroconvulsive therapy? *British Journal of Psychiatry* 1964; 106–109.
8. Scott AIF, Whalley LJ. The onset and rate of the antidepressant effect of electroconvulsive therapy. A neglected topic of research. *British Journal of Psychiatry* 1993; 162:725–732.
9. Parker G, Tennant C, Blignault I. Predicting improvements in patients with non-endogenous depression. *British Journal of Psychiatry* 1985; 146:132–139.
10. Parker G, Blignault I. Psychosocial predictors of outcome in subjects with untreated depressive disorder. *Journal of Affective Disorders* 1985; 8:73–81.
11. Quitkin FM, Stewart JW, McGrath PJ *et al.* Further

- evidence that a placebo response to antidepressants can be identified. *American Journal of Psychiatry* 1993; 150:566–570.
12. Quitkin FM, McGrath PJ, Stewart JW *et al.* Chronological milestones to guide drug change. *Archives of General Psychiatry* 1996; 53:785–792.
 13. Parker G, Roy K, Hadzi-Pavlovic D *et al.* Sub-typing depression by clinical features: The Australasian data base. 1999. *Acta Psychiatrica Scandinavica* 2000 (in press).
 14. Hamilton M. Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology* 1967; 6:278–296.
 15. Parker G, Roussos J, Hadzi-Pavlovic D, Wilhelm K, Mitchell P, Austin M-P. Plumbing the depths: some problems in quantifying depression severity. *Journal of Affective Disorders* 1997; 42:49–58.
 16. Priebe S and Gruyters T. The importance of the first three days: predictors of treatment outcome in depressed in-patients. *British Journal of Clinical Psychology* 1995; 34:229–236.