International Journal of Neuropsychopharmacology (2008), **11**, 15–25. Copyright © 2007 CINP
doi:10.1017/S1461145707007663

ARTICLE

CINP

A pilot effectiveness study: placebo-controlled trial of adjunctive L-triiodothyronine (T3) used to accelerate and potentiate the antidepressant response

Michael Posternak[1], Scott Novak[2], Robert Stern[3], James Hennessey[4], Russell Joffe[5], Arthur Prange Jr.[6] and Mark Zimmerman[7]

[1] *Depression Clinical and Research Program, Massachusetts General Hospital, Boston, MA, USA*
[2] *Division of Health, Social, and Economic Research, Triangle Institute International, Research Triangle Park, NC, USA*
[3] *Department of Neurology, Boston University School of Medicine, Boston, MA, USA*
[4] *Department of Endocrinology, Brown University School of Medicine, Providence, RI, USA*
[5] *Department of Psychiatry, UMDNJ–New Jersey Medical School, Newark, NJ, USA*
[6] *Department of Psychiatry, University of North Carolina School of Medicine, Chapel Hill, NC, USA*
[7] *Department of Psychiatry, Brown University School of Medicine, Providence, RI, USA*

Abstract

The aim was to evaluate whether adjunctive T3 can help accelerate the antidepressant response and improve overall outcomes when used under naturalistic conditions. Fifty consecutive psychiatric outpatients diagnosed with major depressive disorder who were initiated on antidepressant therapy were randomized to receive adjunctive T3 or placebo in a double-blind manner over the course of 6 wk. There were no restrictions placed on the selection of antidepressant agent, dosing, ancillary medications, or psychotherapy, and there were few exclusion criteria. A positive response was defined as a $\geqslant 50\%$ reduction in Montgomery–Asberg Depression Rating Scale scores. Response rates were higher for the adjunctive T3 cohort compared to the adjunctive placebo cohort after 1 wk (45% vs. 24%) and 2 wk (57% vs. 33%) of treatment. The likelihood of experiencing a positive response at any point over the 6-wk trial was 4.5 times greater in the adjunctive T3 cohort (95% CI 1.3–15.7). The study provides preliminary evidence that T3 can successfully be used in clinical practice to accelerate the antidepressant response and improve overall outcomes. The effectiveness model may be an untapped mechanism for evaluating the value of psychopharmacological agents.

Received 6 February 2006; Reviewed 3 April 2006; Revised 15 December 2006; Accepted 15 January 2007; First published online 13 March 2007

Key words: Acceleration, antidepressant, depression, effectiveness, thyroid hormone.

Introduction

Recent reports in both scientific journals and the media have questioned whether the benefits of antidepressant medications have been exaggerated over the years (Fisher and Greenberg, 1997; Goleman, 1995; Horgan, 1998; Kirsch and Sapirstein, 1999; Zimmerman et al., 2002). It has been estimated, for example, that only half of all antidepressant efficacy trials yield positive results (Khan et al., 2002), while negative studies often go unpublished. Even in the positive studies that have been published, the benefits of antidepressant somatic therapy appear to be only modestly better than placebo (Fisher and Greenberg, 1997; Kirsch and Sapirstein, 1999). If only modest results are achieved in highly selected populations conducted under rigorous conditions, how well can these medications be expected to perform in the real world? The differences that exist between findings from controlled research (efficacy) and treatment under naturalistic conditions (effectiveness) has been termed the efficacy-effectiveness gap. The importance to the field in bridging this gap has been well elucidated (Bauer et al., 2001; Wells, 1999), but to date little progress has been made. Standard placebo-controlled trials have rarely been conducted in naturalistic settings, perhaps

Address for correspondence: Professor M. Posternak, 50 Staniford St, Suite 401, Boston, MA, 02114, USA.
Tel.: 617-724-1206 *Fax*: 617-724-3028
E-mail: mposternak@partners.org

because it is widely assumed that drug–placebo differences would be obscured by multiple confounding variables. A second concern might be that if a treatment with proven efficacy cannot be shown to be beneficial in usual clinical practice, a seemingly intractable dilemma would arise as to whether that treatment can be recommended.

The standard methodology used to demonstrate antidepressant efficacy has evolved largely from tradition, however, and there is little empirical evidence suggesting that this methodology is efficient at eliciting drug–placebo differences (Posternak et al., 2002). Increasing attention has been paid recently to the many shortcomings of the traditional design, and at least seven features of the efficacy design may actually serve to obscure drug–placebo differences. First, the majority of antidepressant trials rely on the Hamilton Depression Rating Scale (HDRS). Although the HDRS represented a major advancement at the time it was introduced by standardizing outcome ratings, its shortcomings have been well enumerated: an overemphasis on sleep items, focus on many symptoms peripheral to depression, and absence of items devoted to reversed neurovegetative symptoms (Bagby et al., 2005; Zimmerman et al., 2005). Second, most treatment studies are carried out at multiple sites across the country or world, and training of raters may be inadequate, with few published studies reporting or even establishing inter-rater reliability (Mulsant et al., 2002). Third, outcome ratings are conducted for the most part by research assistants who lack the training, experience, and sophistication that a treating psychiatrist would be expected to have. Poor training and weak reliability increase error variance and could dramatically reduce the ability to detect drug–placebo differences. Fourth, efficacy trials are conducted in artificial settings and are offered as temporary treatment trials. Subjects do not have the opportunity to develop a rapport with a treating psychiatrist. Such dynamics can be expected to lead to higher dropout rates, which poses a significant obstacle to demonstrating drug–placebo differences. Fifth, efficacy trials are often conducted with strong financial incentives, and there may be subtle or overt pressure to recruit subjects quickly. Such an arrangement tends to lead to a relaxation of entry requirements, and the baseline rating scores may get inflated to ensure that subjects meet the minimum symptom severity score requirement (Faries et al., 1999; Robinson and Rickels, 2000). This introduces further error variance. Sixth, dosing regimens tend to be either fixed or restricted, and such restrictions have been shown to reduce drug–placebo differences (Khan et al., 2003). Seventh, clinical trials often require subjects to present for in-person assessments on a weekly basis. These assessments can take 15–30 min or more – a significantly greater amount of contact than occurs in usual clinical practice. This frequent contact has been shown to have a significant therapeutic impact (Posternak and Zimmerman, In Press), which can further reduce drug–placebo separation. Finally, trial investigators who collect outcome ratings are also usually the same ones who inquire about side-effects, and the occurrence of side-effects can 'unblind' randomization (Greenberg et al., 1992). Although unblinding would probably increase rater bias in favour of magnifying drug–placebo differences, this design flaw further undermines the validity and trust in the study's results.

Thus, although it is possible that the benefits of antidepressant medications have been exaggerated as some have suggested, an alternative explanation is that the traditional design used for evaluating antidepressant efficacy may not be an efficient mechanism for separating active medication from placebo. If so, the true benefits of antidepressant therapy may actually be *under*estimated. Conducting randomized trials in naturalistic settings would overcome many of these methodological pitfalls, and although counter to traditional teaching, could in theory demonstrate a superior treatment effect. In addition, of course, effectiveness research enjoys greater ecological validity and generalizability.

A major obstacle to conducting effectiveness research in naturalistic settings is that it can be difficult to recruit subjects into placebo-controlled trials. Placebo-controlled trials in private settings are most likely to succeed if: (1) there is minimal burden placed on patients and clinicians; (2) the study poses minimal deviation from standard clinical practice; (3) patients randomized to placebo receive treatment that approximates usual care; and (4) preliminary evidence exists supporting both the safety and efficacy of the treatment intervention of interest.

An ideal candidate that meets each of these requirements is the use of L-triiodothyronine (T3) as an adjunctive agent to antidepressant therapy for the treatment of major depression. The antidepressant properties of T3 have been recognized for over 30 yr (Earle, 1970). Research has suggested that adjunctive T3 may both hasten the antidepressant response (Altshuler et al., 2001) (i.e. reduce the time to when the antidepressant response occurs), and improve outcomes in patients who have not responded to an initial adequate antidepressant trial (Abraham et al., 2006; Aronson et al., 1996). Of note, however, a recent placebo-controlled study by Appelhof et al.

(2004) found that T3 did not help accelerate the antidepressant response when added to paroxetine, nor did it improve response rates at end-point. As a natural substance, T3 is considered to be one of the safest psychopharmacological agents available. Nevertheless, despite modest empirical support, a favourable side-effect profile, and a generic formulation, adjunctive T3 is rarely used in clinical practice (Byrne and Rothschild, 1997; Chaimowitz et al., 1991; Fredman et al., 2000; Shergill and Katona, 1997). The reasons for the under-utilization of T3 are unclear, but may stem from problems inherent in the T3 research conducted to date. Limitations of most of this research (other than and prior to the study by Appelhof et al.) include small sample sizes (range 4–35 subjects), focus on psychiatric in-patients rather than outpatients, and a paucity of data with the newer generation of antidepressants (Lasser and Baldessarini, 1997).

The goals of the present study were therefore twofold: (1) to evaluate whether the results of the T3 research conducted to date – most of which was performed over 25 yr ago – can be extended to today's practice; and (2) to determine whether drug–placebo differences can be elicited using an effectiveness rather than efficacy trial design, thereby demonstrating the benefits of a somatic intervention as it might be used in usual clinical practice. Design features we implemented in the present study to enhance drug–placebo separation include: (1) using the Montgomery–Asberg Depression Rating Scale (MADRS; Montgomery and Asberg, 1979) as the primary outcome measure, and validating all outcome ratings with a self-rated instrument; (2) having the treating psychiatrist conduct outcome ratings; (3) conducting all assessments at a single site after demonstrating strong inter-rater reliability among all raters; (4) conducting the trial in a naturalistic setting, which we hypothesize will lead to lower dropout rates; (5) allowing flexible dosing schedules; (6) assessing side-effects only after outcome ratings had been collected; and (7) absence of financial incentives.

Method

All subjects were recruited from the Rhode Island Hospital Department of Psychiatry's outpatient practice. This is a fee-for-service practice that functions independently from the Brown University Residency Program. At the time of presentation and prior to meeting their treating clinician, patients were invited to undergo a research diagnostic evaluation as part of the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project (Zimmerman and Mattia, 1999, 2000). This evaluation consists of the Structured Clinical Interview for DSM-IV (SCID; First et al., 1997) and the Structured Interview for DSM-IV Personality Disorders (SIDP-IV; Pfohl et al., 1997), as well as various other clinician- and patient-rated instruments. This evaluation is most often conducted by clinical psychologists who have undergone extensive training. Fifteen subjects treated by either Dr Posternak or Dr Zimmerman who did not participate in the MIDAS project were also recruited. For these individuals, Axis I diagnoses were established using the Psychiatric Diagnostic Screening Questionnaire (PDSQ; Zimmerman and Mattia, 2002), followed by a clinical evaluation by the treating psychiatrist. The presence of borderline personality disorder (BPD) was evaluated in this group using the BPD component of the SIDP-IV. No baseline demographic or clinical differences were found between subjects who did and did not undergo a research diagnostic evaluation.

Eligibility criteria for the study included being aged at least 18 yr and meeting full DSM-IV criteria for major depressive disorder (MDD). Subjects who had unstable cardiac, endocrine, or renal disease, a history of thyroid disease, or an abnormal baseline thyrotropin (normal range 0.3–5.5 uIU/ml), were excluded. Other than medical contraindication to T3 therapy, however, there were no restrictions to participation. Thus, patients diagnosed with bipolar disorder, psychotic features, psychiatric comorbidity, or a history of treatment resistance, were all invited to participate.

The trial was designed as a pilot effectiveness study to establish feasibility and to evaluate whether drug–placebo differences could be elicited using this model. As such, a sample size of 50 subjects was targeted. The study was therefore not powered to find significant differences. With a sample size of 25 subjects per cohort, and an estimated effect size of 0.6 (Altshuler et al., 2001), there was approximately a 55% chance of observing significant differences between groups. The protocol was approved by the Rhode Island Hospital Institutional Review Board and all subjects provided informed, written consent.

Consecutive subjects were recruited at the time an antidepressant medication was initiated. The present study therefore focuses on the ability of T3 to *accelerate* (reduce the delay in time to response) and *potentiate* (improve outcomes at end-point) the antidepressant response, but does not evaluate T3 as an augmentation agent for treatment non-responders. All treatment, except for adjunctive T3 and placebo, was open label and followed usual clinical practice. No restrictions were placed on the selection of antidepressant, dosage, ancillary medications, or psychotherapy. The antidepressant agent was not changed during the course

of the 6-wk trial. All participating subjects were randomized to receive either adjunctive T3 at a dose of 0.025 mg/d or placebo in identically appearing capsules each morning in a double-blind manner over the course of 6 wk. We chose the lower 0.025 mg/d dosage as opposed to the 0.05 mg/d dosage, because this has become more commonly used in prior research. Randomization was accomplished by having the pharmacist pre-sort study pills, and allowing the treating clinician to randomly pick coded vials to give to a study subject at the time of recruitment. The study medication was typically initiated on day 2 rather than day 1 because subjects were instructed not to take the study pill until the baseline thyrotropin was confirmed to be within the normal range. Compliance was not formally monitored.

The primary outcome measure was the MADRS. We chose the MADRS over other instruments because it is relatively brief, and may be more sensitive to change than the HDRS. Inter-rater reliability for the MADRS and was established in 30 joint interviews. The intra-class correlation coefficient for these interviews was 0.96. The self-rated Clinically Useful Depression Outcome Scale (CUDOS; Zimmerman et al., 2004b) was used as a secondary outcome measure. The CUDOS was chosen because it is brief, it is directly tied to DSM-IV, it assesses reversed neurovegetative symptoms, and it is has a validated cut-off for remission (a CUDOS score of <20) (Zimmerman et al., 2004b). Antidepressant treatment history was elicited using the Treatment Response to Antidepressant Questionnaire (TRAQ). The TRAQ is a semi-structured instrument developed by our group with demonstrated reliability (Posternak et al., 2004) and validity (Posternak and Zimmerman, 2003).

Outcome ratings were collected by the treating psychiatrist at baseline, weeks 1, 2, 3, and 6. Ratings for weeks 1-3 focused on the putative ability of T3 to accelerate the antidepressant response, while week 6 ratings focused on the ability of T3 to potentiate the antidepressant response. In-person follow-up appointments were typically scheduled at week 3 and week 6 (although there were no restrictions on this), consistent with our prior research (Posternak and Zimmerman, 2001). Because participating subjects were not reimbursed for their participation, it was deemed overly burdensome to require them to present for weekly visits. Therefore, MADRS and CUDOS ratings for weeks 1 and 2 were usually conducted by telephone. Telephone ratings have been demonstrated to yield reliable and valid results (Mundt et al., 2006).

Side-effects to T3 therapy were assessed at week 3 and week 6 using a standardized hyperthyroid checklist (Braverman and Utiger, 2000), which was filled out after all other outcome ratings had been obtained. At week 6, subjects and clinicians were asked to guess on a 5-point scale whether they were 'almost sure' or believed they 'probably' had received T3 (or placebo), or whether they were 'not sure' which study pill they had received.

Our two principal hypotheses were that (1) adjunctive T3 will accelerate the antidepressant response from baseline to week 3, and (2) adjunctive T3 will improve overall response rates at end-point. To test both hypotheses, we conducted categorical (i.e. whether subjects achieved a $\geqslant 50\%$ reduction in baseline MADRS scores) and dimensional (i.e. mean change) analyses. We used the Generalized Estimating Equation (GEE) approach to evaluate the first hypothesis regarding the ability of T3 to accelerate the antidepressant response during the first 3 wk of treatment (Diggle et al., 1994; Liang and Zeger, 1986; Zeger and Liang, 1986). This method was chosen because it adjusts the variance components of the parameter estimates, which can become underestimated in the presence of correlated data. This is particularly relevant for longitudinal data where the within-subject correlations are increased due to repeated measurements collected on the same set of individuals over time. However, we also report the score χ^2 test statistic for each statistical test (e.g. Z statistic), which has been shown to be more conservative than those based on the empirical and model-based standard errors, and is preferred for small samples (Stokes et al., 2000).

The estimates of the standard errors, which are model-based, were derived from unstructured working correlational matrices given the relatively few number of data-points per subject. The general model specification to test the key hypothesis is as follows:

$$\log\left(\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)}\right)=\mu+\alpha*\text{Treatment}+\beta*\text{Weeks}+\delta*(\text{Treatment}*\text{Weeks}),$$

where α is the main effect of Treatment status ($0=$control, $1=$T3), β is the main effect of time, as measured in weeks (1, 2 and 3), and δ is the interaction between Treatment and Weeks. We also reformulated this model to accommodate a continuous distribution based on the raw scores of the MADRS for each week. The baseline wave was also included in this model, which consequently included four waves of data and a 4-level time-varying covariate for weeks (0, 1, 2, and 3).

Tests of the hypotheses concerning the main effect of the intervention within each week and at end-point were conducted using the standard logistic regression model for the binary outcomes (Hosmer and

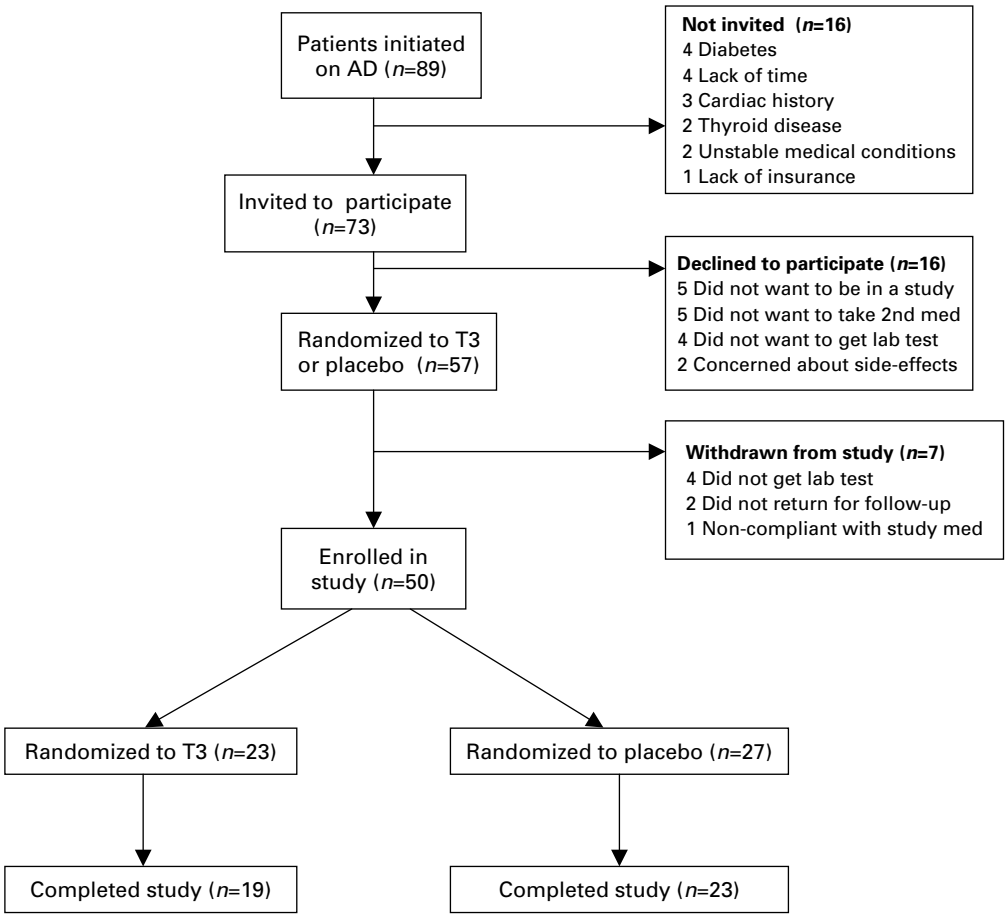

Figure 1. Flow chart of enrolment of subjects.

Lemeshow, 2000). Analysis of covariance (ANCOVA) was used for the dimensional analyses to estimate difference in treatment effects with the baseline MADRS scores as covariates, using the last observation carried forward (LOCF). Remission from depression was defined as an end-point MADRS score of $\leqslant 10$ (Zimmerman et al., 2004a).

Results

Recruitment and baseline characteristics

Eighty-nine subjects with MDD were initiated on an antidepressant medication during the study period. Of these, 16 subjects were excluded (most due to medical comorbidity), and 16 others declined to participate (see Figure 1). The remaining 57 subjects were randomized to a study medication. Of these, seven were withdrawn or dropped out prior to the week 1 follow-up visit. No differences were found in baseline features between subjects who did and did not participate. Of the 50 subjects who participated in the trial, 23 were randomized to adjunctive T3 and 27 to placebo. There were no statistically significant differences in any of the baseline demographic or clinical features between these two cohorts (Table 1).

Treatments received

Selective serotonin reuptake inhibitors (SSRIs) constituted the majority of antidepressant prescriptions ($n=26$, 52%), followed by bupropion ($n=8$, 16%), venlafaxine ($n=7$, 14%), and mirtazapine ($n=4$, 8%) (Table 2). All subjects except one received what is generally considered a minimum adequate dosage (Sackeim et al., 1990) (that one subject had responded to 100 mg/d nefazodone, and the dosage was not increased further because she also experienced side-effects). Thirty-two (64%) subjects received one or more ancillary medications during the course of their treatment trial: 19 (38%) received a sedative-hypnotic, 16 (32%) an anxiolytic medication, three (6%) an antipsychotic, one (2%) a mood stabilizer, and one (2%) a stimulant. Ancillary medications were initiated

Table 1. Baseline demographic and clinical features of T3 cohort and control group[a]

	AD + T3 ($n = 23$)	AD + placebo ($n = 27$)
Female, *n* (%)	13 (57)	19 (70)
Age (yr), mean ± S.D.	40 ± 9.4	36 ± 11.4
Race: white, *n* (%)	21 (91)	24 (89)
Marital status, *n* (%)		
Single	7 (30)	10 (37)
Married/living together	11 (48)	11 (41)
Divorced/separated/widowed	5 (22)	6 (22)
Education, *n* (%)		
Less than high-school diploma	1 (4)	1 (4)
High-school graduate or GED	14 (51)	19 (70)
College or postgraduate degree	8 (40)	7 (26)
Antidepressant status		
Newly initiated antidepressant	15 (65)	18 (67)
Switch following non-response	6 (26)	5 (19)
Switch following relapse	2 (9)	4 (15)
Episode duration (months)	51 ± 80	27 ± 37
Depression subtype		
Unipolar MDD	20 (87)	21 (78)
Bipolar disorder, I or II	3 (13)	3 (11)
MDD with psychotic features	0 (0)	3 (11)
Comorbidity		
No psychiatric comorbidity	6 (26)	3 (11)
Panic disorder ± agoraphobia	5 (22)	8 (30)
Post-traumatic stress disorder	5 (22)	3 (11)
Obsessive–compulsive disorder	1 (4)	2 (7)
Generalized anxiety disorder	10 (44)	11 (41)
Social phobia	7 (30)	4 (15)
Dysthymia	1 (4)	1 (4)
Eating disorder	1 (4)	4 (15)
Alcohol or drug abuse	2 (9)	6 (22)
Borderline personality disorder	3 (13)	4 (15)
Antidepressant treatment history		
No prior adequate trials	7 (30)	7 (26)
1 failed adequate trial	4 (17)	6 (22)
2 failed adequate trials	1 (4)	2 (7)
⩾3 failed adequate trials	3 (13)	2 (7)
⩾1 positive trials	1 (4)	5 (19)
Some pos. and some neg. trials	7 (30)	5 (19)
Baseline MADRS	29.3 ± 7.6	30.3 ± 8.7
Baseline CUDOS	36.9 ± 9.8	39.7 ± 8.6
Baseline thyrotropin level	1.5 ± 0.7	1.5 ± 0.8

GED, General equivalency diploma; MDD, major depressive disorder; MADRS, Montgomery–Asberg Depression Rating Scale; CUDOS, Clinically Useful Depression Outcome Scale.

[a] No statistically significant differences in baseline features between the two cohorts.

(rather than continued) for 7/23 (30%) subjects randomized to T3 compared to 15/27 (56%) subjects randomized to placebo ($\chi^2 = 3.2$, d.f. = 1, $p = 0.07$, n.s.). Twenty-five (50%) subjects received at least one psychotherapy session during the course of their treatment trial. Eleven (22%) subjects initiated psychotherapy and 14 (28%) were continued in psychotherapy. Subjects randomized to placebo received a greater number of psychotherapy sessions (1.9 ± 2.1) than those randomized to T3 (0.6 ± 0.8) ($t = 2.7$, d.f. = 35.4, $p = 0.007$). There were no other statistically significant differences between the two cohorts in treatments received.

Forty-two of the 50 (84%) subjects completed the 6-wk trial: 19/23 (83%) subjects receiving adjunctive T3 and 23/27 (85%) subjects receiving placebo completed the trial. As hypothesized, the retention rate for patients in the present study was significantly greater ($\chi^2 = 9.59$, d.f. = 1, $p = 0.002$) than the 62.7% rate reported in a recent meta-analysis of antidepressant trials (Posternak and Zimmerman, 2005).

Treatment response

Acceleration

Response rates were non-significantly higher according to the MADRS during each of the first 3 wk of treatment in the adjunctive T3 cohort compared to the control group: week 1, 45% vs. 24%; week 2, 57% vs. 33%; and week 3, 43% vs. 24%. The GEE model evaluating dichotomous outcomes of response for weeks 1–3 revealed no significant main effects for condition ($z = 1.47$, $p = 0.14$), week ($z = 0.82$, $p = 0.41$), or the interaction ($z = 0.02$; $p = 0.99$). An analysis evaluating continuous outcomes (Figure 2) revealed a significant effect favouring T3 over the first 3 wk of treatment ($z = 2.0$, $p < 0.05$). The mean MADRS scores for the T3 cohort over weeks 1–3 were 18.4 (S.D. = 10.3), 14.0 (S.D. = 9.4), and 17.2 (S.D. = 13.0). The mean MADRS scores for the adjunctive placebo cohort during the same time-frame were 22.3 (S.D. = 12.5), 23.5 (S.D. = 14.2), and 21.1 (S.D. = 13.6), respectively. Differences in MADRS scores reached statistical significance only at week 2.

When outcomes were examined using the self-rated CUDOS instrument, a similar pattern of improvement was observed, although these differences did not reach statistical significance (Figure 3). In both the self-rated and clinician-rated instruments, both groups demonstrated a slight increase in depression severity scores at week 3. We are unclear as to why this might be, but suspect that it is due to random variation that sometimes occurs when small sample sizes are employed.

Table 2. Antidepressant medications, dosages, and ancillary treatments in T3 cohort and control group

	AD + T3 ($n = 23$)	AD + placebo ($n = 27$)
Antidepressant, *n* (median dosage for week 3/week 6)		
Fluoxetine	3 (20/40 mg)	9 (20/20 mg)
Sertraline	6 (50/50 mg)	2 (50/75 mg)
Escitalopram	3 (10/15 mg)	3 (10/20 mg)
Venlafaxine XR	5 (75/150 mg)	2 (150/150 mg)
Mirtazapine	1 (30/45 mg)	3 (15/30 mg)
Bupropion	3 (200/200 mg)	5 (200/350 mg)
Phenelzine	0 (0/0 mg)	1 (45/45 mg)
Nefazodone	1 (600/600 mg)	1 (100/100 mg)
Amitriptyline	1 (125/150 mg)	0 (0/0 mg)
Imipramine	0 (0 mg)	1 (150/150 mg)
Ancillary psychiatric medications		
Sedative/hypnotics	8 (27)	11 (31)
Anxiolytics	7 (18)	9 (31)
Mood stabilizers	0 (0)	1 (6)
Stimulants	0 (0)	1 (6)
Antipsychotics	0 (0)	3 (19)
Initiated $\geqslant 1$ ancillary medications	6 (26)	12 (44)
Psychotherapy		
Received 1 or more therapy sessions	9 (39)	16 (59)
Initiated psychotherapy	5 (22)	6 (22)
Continued in psychotherapy	4 (17)	10 (37)
Number of therapy sessions[a] (mean ± S.D.)	0.6 ± 0.8	1.9 ± 2.1

[a] $t = 2.7$, d.f. = 35.4, $p = 0.007$.

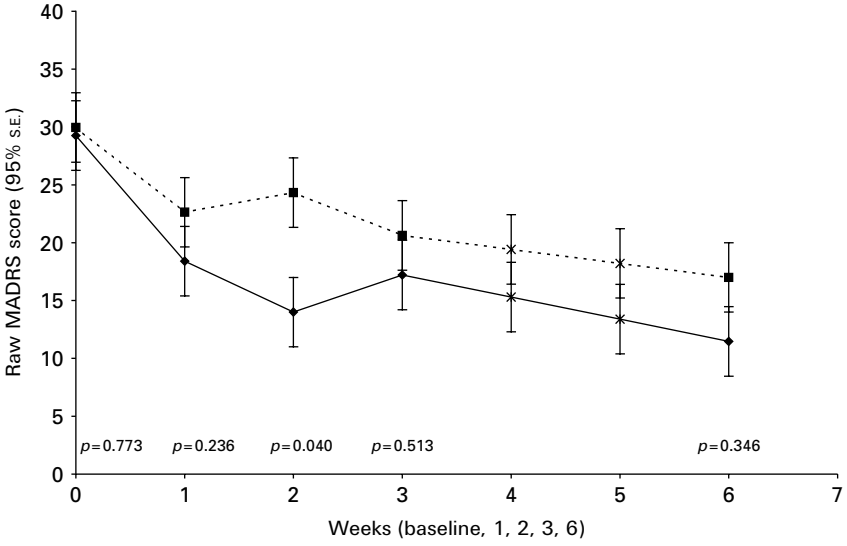

Figure 2. Time-course of improvement on adjunctive T3 (–◆–) and placebo (- -■- -). *p* values for ANCOVA tests within weeks 1, 2, 3, and 6 control for baseline values of MADRS (weeks 4 and 5 imputed).

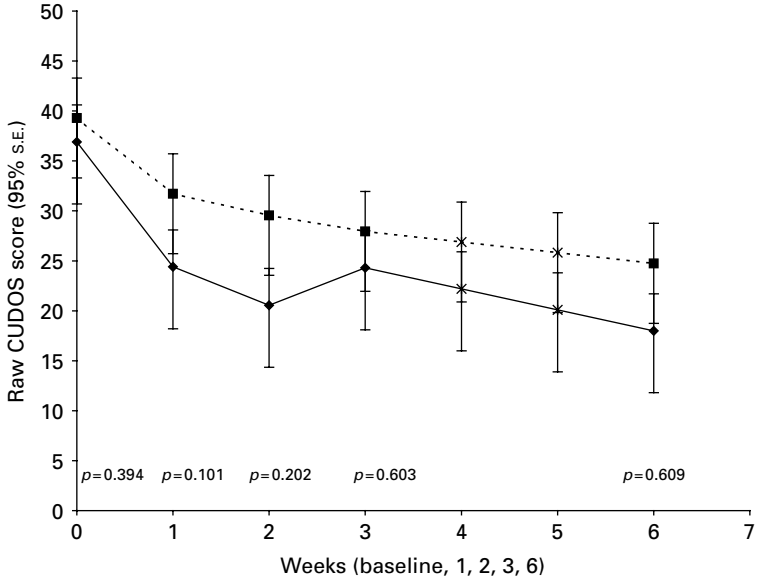

Figure 3. Time-course of improvement on adjunctive T3 (–◆–) and placebo (- -■- -). *p* values for ANCOVA tests within weeks 1, 2, 3, and 6 control for baseline values of CUDOS (weeks 4 and 5 imputed).

Potentiation

At end-point, with LOCF, response rates were higher for subjects receiving T3 than placebo (61% vs. 52%), although this difference was not statistically significant (OR 1.44, 95% CI 0.46–4.46, $p=0.52$). Remission rates were also numerically higher according to both the MADRS (48% vs. 37%) and CUDOS (55% vs. 33%), although again these differences did not reach statistical significance ($p=0.44$ and $p=0.14$, respectively).

Side-effects

Side-effects were assessed using an 11-item self-rated checklist assessing symptoms consistent with hyperthyroidism (Table 3). For 10 out of 11 of these side-effects, incidence rates were numerically higher in the cohort receiving placebo. The only statistically significant difference occurred for reports of nervousness, which was significantly more common in the cohort receiving placebo (10/22, 45%) than in the T3 cohort (2/19, 11%) ($\chi^2=6.0$, d.f.$=1$, $p=0.01$). Other studies (e.g. Appelhof et al., 2004) have found that T3 does induce a consistent and predictable side-effect profile. Our inability to elicit side-effects may again be a function of the small sample size employed.

Blinding

At the conclusion of the trial, subjects and clinicians were asked to make a guess as to randomization assignment along with degree of conviction. Two-thirds of all subjects reported that they were not sure which assignment they had received. In the placebo cohort, 2/23 (9%) subjects thought or were almost sure they were receiving T3, while 3/23 (13%) thought they were receiving placebo. In the T3 cohort, 5/19 (26%) thought or were almost sure they were receiving T3, while 4/19 (21%) thought they were receiving placebo.

From the clinician standpoint, of 23 subjects randomized to placebo, clinicians reported thinking that six (26%) were receiving T3 and judged 2/23 (9%) to be receiving placebo. In 20 subjects receiving T3, clinicians guessed correctly in seven (35%) instances and incorrectly in four (20%) instances. In the remaining instances, clinicians were unsure as to randomization.

Discussion

Traditionally, the therapeutic effects of a medication are established under highly controlled conditions de signed to maximize the likelihood of eliciting drug–placebo differences. In antidepressant trials, subjects with mild depression, a history of treatment resistance, or psychiatric comorbidity are routinely excluded. Ancillary treatments are usually prohibited or restricted in order to eliminate potential confounding variables. Once efficacy is established under these conditions, effectiveness is inferred for patients treated in the real world. This inference cannot be assured, however, because treatment conditions are distinct and patients in the real world may differ dramatically

Table 3. Side-effects reported in the adjunctive T3 and placebo cohorts

	T3 ($n=19$)	Placebo ($n=22$)
Nervousness[a]	2 (11%)	10 (46%)
Fatigue	4 (21%)	7 (32%)
Weakness	3 (16%)	4 (18%)
Increased sweating	2 (11%)	6 (27%)
Heat intolerance	4 (21%)	1 (5%)
Tremor	0 (0%)	2 (9%)
Hyperactivity	1 (5%)	2 (9%)
Palpitations	2 (11%)	3 (14%)
Appetite increase	2 (11%)	3 (14%)
Weight decrease	2 (11%)	7 (32%)
Menstrual disturbances[b]	0 (0%)	1 (4%)

[a] Statistically significant ($\chi^2=6.0$, d.f. $=1$, $p=0.01$).
[b] Based on subsample of women for T3 ($n=9$) and placebo ($n=14$).

from those who participate in treatment studies (Zimmerman et al., 2002). Ideally, the effectiveness of all psychopharmacological agents could be established under naturalistic conditions, but conducting controlled trials in real-world settings presents multiple pragmatic obstacles. The lack of research under naturalistic conditions has led to questions as to the true effectiveness of psychopharmacological agents in actual clinical practice (Zimmerman et al., 2002).

Psychotherapy researchers have already begun to establish that findings from controlled psychotherapy research can be exported to naturalistic settings (Franklin et al., 2000; Persons and Silverschatz, 1998; Wade et al., 1998), but we are not aware of any comparable attempts in psychopharmacology. A placebo-controlled augmentation trial with T3 seemed to be an ideal starting point to help bridge the efficacy-effectiveness gap. T3 is well-tolerated, safe, and has preliminary empirical support. Using an augmentation paradigm, subjects who were randomized to placebo received the same treatment they would have had they not participated in the study (except for the placebo pill). To the best of our knowledge, the present study is first to directly evaluate whether the specific benefits of a psychopharmacological agent can be demonstrated while used under almost entirely naturalistic conditions. Such conditions also potentially allow for a much richer evaluation of the study population, including rigorous assessments of comorbidity and treatment history. Larger studies could utilize such data to perform sub-analyses of predictors of response that have heretofore rarely been attempted.

Even without preferentially recruiting subjects who might be more likely to respond to T3 or placing restrictions on ancillary treatment, our results are suggestive that adjunctive T3 may help accelerate the antidepressant response in clinical practice, and perhaps improve overall outcomes. The benefits of T3 were most apparent early in treatment, although separation from the control group persisted to end-point. These results must be viewed cautiously, however, since the present study employed a relatively small sample size and many of the differences did not reach statistical significance. The present study must therefore be viewed only as pilot in nature. We also can not rule out that baseline differences – such as less psychiatric comorbidity in the T3 group – might have been at least partially responsible for the drug–placebo separation.

Nevertheless, our ability elicit even some drug–placebo separation under naturalistic conditions with only modest sample sizes and without even attempting to control for potential confounding factors is encouraging on three counts. First, it provides further evidence that T3 can be used to help accelerate the antidepressant response, and may improve response rates at the conclusion of a 6-week trial. Second, our results provide preliminary evidence that findings from controlled research may be able to be replicated when conducted under naturalistic conditions. Third, our study raises the possibility that naturalistic settings may offer an untapped paradigm to evaluate drug efficacy. Innovations that we made to attempt to overcome potential obstacles present in traditional research were: utilizing the MADRS as opposed to the HDRS, having the treating psychiatrist conduct all outcome ratings after demonstrating strong inter-rater reliability, conducting the study entirely at one site, allowing for flexible dosing, reducing the amount of contact with the research clinician (to minimize the non-specific therapeutic effects of such interactions), and a lack of financial incentives. Although it is impossible to determine what impact these factors had on outcomes, the present study at least demonstrated the feasibility of conducting placebo-controlled research in clinical settings, while obtaining high recruitment and retention rates.

Considering the enormous costs and consequences of employing a potentially inefficient study design to evaluate drug efficacy, it is surprising that more effort has not gone into studying the impact of various design features. In addition to overcoming many of the pitfalls that are present in traditional efficacy studies, controlled research in naturalistic settings is undoubtedly the best way to gauge the true value of a psychopharmacological intervention.

Further research is warranted to confirm that T3 can help accelerate the antidepressant response in clinical practice and improve overall response rates. Even a modest augmentation in response rates could have an enormous public health impact. Positive results could also help instil confidence in using an effectiveness model for evaluating the benefits of other psychopharmacological agents.

Acknowledgements

The investigators gratefully acknowledge all the patients who participated in this study, and who did so without any tangible benefit to themselves other than to help advance the field.

Statement of Interest

None.

References

Abraham G, Milev R, Stuart JL (2006). T3 augmentation of SSRI resistant depression. *Journal of Affective Disorders 91*, 211–215.

Altshuler LL, Bauer M, Frye MA, Gitlin MJ, Mintz J, Szuba MP, Leight KL, Whybrow PC (2001). Does thyroid supplementation accelerate tricyclic antidepressant response? A review and meta-analysis of the literature. *American Journal of Psychiatry 158*, 1617–1622.

Appelhof BC, Brouwer JP, Dyck RV, Fliers E, Hoogendijk WJG, Huyser J, Schene AH, Tijssen JG, Wiersinga WM (2004). Triiodothyronine addition to paroxetine in the treatment of mjoar depressive disorder. *Journal of Clinical Endocrinology and Metabolism 89*, 6271–6276.

Aronson R, Offman HJ, Joffe RT, Naylor CD (1996). Triiodothryonine augmentation in the treatment of refractory depression. A meta-analysis. *Archives of General Psychiatry 53*, 842–848.

Bagby RM, Ryder AG, Schuller DR, Marshall MB (2005). The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *American Journal of Psychiatry 161*, 2163–2177.

Bauer MS, Williford WO, Dawson EE, Akiskal HS, Altshuler L, Fye C, Gelenberg A, Glick H, Kinosian B, Sajatovic M (2001). Principles of effectiveness trials and their implementation in VA Cooperative Study #430: Reducing the efficacy-effectiveness gap in bipolar disorder. *Journal of Affective Disorders 67*, 61–78.

Braverman J, Utiger T (2000). Thyrotoxicosis. In: Werner T, Ingbar C (Eds.), *The Thyroid* (vol. 8, pp. 515–517). Philadelphia: Lippincott, Williams & Wilkins.

Byrne S, Rothschild AJ (1997). Psychiatrists' responses to failure of maintenance therapy with antidepressants. *Psychiatric Services 48*, 835–837.

Chaimowitz GA, Links PS, Padgett RW, Carr AC (1991). Treatment-resistant depression: a survey of practice habits of Canadian psychiatrists. *Canadian Journal of Psychiatry* 36, 353–356.

Diggle PJ, Liang KY, Zeger SL (1994). *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

Earle BV (1970). Thyroid hormone and tricyclic antidepressants in resistant depression. *American Journal of Psychiatry 126*, 143–145.

Faries DE, Heilegenstein JH, Tollefson GD, Potter WZ (2001). Double-blind variable placebo lead-in period: results from two antidepressant clinical trials. *Journal of Clinical Psychopharmacology 21*, 561–568.

First MB, Spitzer RL, Williams JBW, Gibbon M (1997). *Structured Clinical Interview for DSM-IV* (*SCID*). Washington, DC: American Psychiatric Association.

Fisher S, Greenberg RP (Eds.). (1997). *From Placebo to Panacea*. New York: John Wiley & Sons.

Franklin ME, Abramowitz JS, Levitt JT, Kozak MJ, Foa EB (2000). Effectiveness of exposure and ritual prevention of obsessive-compulsive disorder: randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology 68*, 594–602.

Fredman SJ, Fava M, Kienke AS, White CN, Nierenberg AA, Rosenbaum JF (2000). Partial response, nonresponse, and relapse with selective serotonin reuptake inhibitors in major depression: a survey of current 'next-step' practices. *Journal of Clinical Psychiatry 61*, 403–408.

Goleman D (1995). Psychologists dispute value of antidepressants. *New York Times*, 29 November 1995, p. C1910.

Greenberg RP, Bornstein RF, Greenberg MD, Fisher S (1992). A meta-analysis of antidepressant outcome under 'blinder' conditions. *Journal of Consulting and Clinical Psychology 60*, 664–669.

Horgan J (1998). Science triumphant? Not so fast. *New York Times*, 19 January 1998, p. A1917.

Hosmer D, Lemeshow S (2000). *Applied Logistic Regression*, vol. 2. San Francisco, CA: John Wiley and Sons.

Khan A, Khan S, Brown WA (2002). Are placebo controls necessary to test new antidepressants and anxiolytics? *International Journal of Neuropsychopharmacology 5*, 193–197.

Khan A, Khan SR, Walens G, Kolts R, Giller EL (2003). 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 28*, 552–557.

Kirsch I, Sapirstein G (1999). Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medications. In: Kirsch I (Ed.), *How Expectancies Shape Experience* (pp. 303–320). Washington, DC: American Psychological Association.

Lasser RA, Baldessarini RJ (1997). Thyroid hormones in depressive disorders: a reappraisal of clinical utility. *Harvard Review of Psychiatry 4*, 291–305.

Liang KY, Zeger SL (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Montgomery SA, Asberg M (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry 134*, 382–389.

Mulsant BH, Kastango KB, Rosen J, Stone RA, Mazumdar S, Pollock BG (2002). Interrater reliability in clinical trials of depressive disorders. *American Journal of Psychiatry 159*, 1598–1600.

Mundt JC, Katzelnick DJ, Kennedy SH, Eisfeld BS, Bouffard BB, Greist HJ (2006). Validation of an IVRS version of the MADRS. *Journal of Psychiatric Research 40*, 243–246.

Persons JB, Silverschatz G (1998). Are results of randomized controlled trials useful to psychotherapists? *Journal of Consulting and Clinical Psychology 66*, 126–135.

Pfohl B, Blum N, Zimmerman M (1997). *Structured Interview for DSM-IV Personality*. Washington, DC: American Psychiatric Press Inc.

Posternak MA, Young D, Sheeran T, Chelminski I, Franklin CL, Zimmerman M (2004). Assessing past treatment history: the test-retest reliability of the Treatment Response to Antidepressant Questionnaire. *Journal of Nervous and Mental Disease 192*, 95–102.

Posternak MA, Zimmerman M (2001). Switching versus augmentation: a prospective, naturalistic comparison in depressed, treatment-resistant patients. *Journal of Clinical Psychiatry 62*, 135–142.

Posternak MA, Zimmerman M (2003). How accurate are patients in reporting their antidepressant treatment history? *Journal of Affective Disorders 75*, 115–124.

Posternak MA, Zimmerman M (2005). Is there a delay in the antidepressant effect? A meta-analsyis. *Journal of Clinical Psychiatry 66*, 148–158.

Posternak MA, Zimmerman M (In Press). Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials. *British Journal of Psychiatry*.

Posternak MA, Zimmerman M, Keitner GI, Miller IW (2002). A reevaluation of the exclusion criteria used in antidepressant efficacy trials. *American Journal of Psychiatry 159*, 191–200.

Robinson DS, Rickels K (2000). Concerns about clinical drug trials. *Journal of Clinical Psychopharmacology 6*, 593–596.

Sackeim HA, Prudic J, Devanand DP, Decina P, Kerr B, Malitz S (1990). The impact of medication resistance and continuation pharmacotherapy on relapse following response to electroconvulsive therapy in major depression. *Journal of Clinical Psychopharmacology 10*, 96–104.

Shergill SS, Katona CLE (1997). Pharmacological choices after one antidepressant fails: a survey of UK psychiatrists. *Journal of Affective Disorders 43*, 19–25.

Stokes ME, Davis CS, Kocki GG (2000). *Categorical Data Analysis using the SAS System* (2nd edn.). Cary, NC: SAS Institute.

Wade WA, Treat TA, Stuart GL (1998). Transporting an empirically supported treatment for panic diosrder to a service clinic setting: a benchmarking strategy. *Journal of Consulting and Clinical Psychology 66*, 231–239.

Wells KB (1999). Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. *American Journal of Psychiatry 156*, 5–10.

Zeger S, Liang L (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Zimmerman M, Mattia JI (1999). Psychiatric diagnosis in clinical practice: Is comorbidity being missed? *Comparative Psychiatry 40*, 182–191.

Zimmerman M, Mattia JI (2000). Principal and additional DSM-IV disorders for which outpatients seek treatment. *Psychiatric Services 51*, 1299–1304.

Zimmerman M, Mattia JI (2002). A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. *Archives of General Psychiatry 58*, 787–794.

Zimmerman M, Mattia JI, Posternak MA (2002). Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *American Journal of Psychiatry 159*, 469–473.

Zimmerman M, Posternak MA, Chelminski I (2004a). Defining remission on the Montgomery–Asberg Depression Rating Scale. *Journal of Clinical Psychiatry 65*, 163–168.

Zimmerman M, Posternak MA, Chelminski I (2004b). Using a self-report depression scale to identify remission in depressed outpatients. *American Journal of Psychiatry 161*, 1911–1913.

Zimmerman M, Posternak MA, Chelminski I (2005). Is it time to replace the Hamilton Depression Rating Scale as the primary outcome measure in treatment studies of depression? *Journal of Clinical Psychopharmacology 25*, 105–110.