COMMENTS FROM 3 *JAMA* REVIEWERS

AND

SUGGESTED REVISIONS TO BE MADE BEFORE SUBMITTING TO

AMERICAN JOURNAL OF PSYCHIATRY

Title: Efficacy of paroxetine but not imipramine in the treatment of adolescent major depression: A randomized controlled trial.

Authors: Keller et al.

SCIENTIFIC THERAPEUTICS
INFORMATION, INC

505 Morris Avenue
Springfield, New Jersey
07081

(973) 376-3635 telephone
(973) 376-0611 fax 1
(973) 376-3367 fax 2

http: \\www.stimedinfo.com
E-mail: staff@stimedinfo.com

June 12, 2000

Martin B. Keller, MD
Department of Psychiatry and Human Behavior
Brown University School of Medicine
345 Blackstone Boulevard
Providence, RI 02906

Dear Dr Keller:

We are pleased to enclose all of the necessary materials for you to submit your manuscript, "Efficacy of Paroxetine but Not Imipramine in the Treatment of Adolescent Major Depression: A Randomized, Controlled Trial," to the *Journal of the American Academy of Child and Adolescent Psychiatry*.

Please find enclosed the following items:

- Five copies of the manuscript (submit four to the journal; keep one for your files)
- One set of glossy prints of the figures (submit to the journal)
- A draft cover letter to Dr Dulcan, editor of JAACAP (please retype on your letterhead and revise as you like)
- A diskette containing your manuscript in Microsoft Word 97 format for your use in the event revisions are needed
- Your disclosure information (submit to the journal).

On behalf of Sally Laden, it has been a pleasure working with you on this project. Please keep us apprised of the status of the paper. If revisions are required, we will be happy to assist you.

Thank you for your cooperation, and please do not hesitate to contact us with questions.

Kindest Regards,

Erika Dankovits
Associate Copy Editor

cc: J Romankiewicz, M Philips, S Laden, J McCafferty, B Brand, .1301

encl.

Running head: Paroxetine Treatment of Adolescent Depression

Corresponding author: Martin B. Keller, MD, Department of Psychiatry and Human Behavior, Brown University School of Medicine, 345 Blackstone Boulevard, Providence, RI 02906, telephone: (401) 455-6430, fax: (401) 450-6441, E-mail: kelly_griffin@brown.edu

Statistical expert: Rosemary Oakes, MS

Word count: 6,522

Reviewer #1:

This paper reports on a large double-blind trial of paroxetine and imipramine versus placebo. There are several problems with this study as follows:

1. The major finding of this study was the high placebo response rate, nearly 50%. Paroxetine produced only a 20% higher response rate than placebo and then only on some but not all of the scales used. The parent and patient self-report scales did not show a difference. The superiority of paroxetine over placebo came because of the numbers studied rather than the effect size of the drug. Readers of this paper might receive the wrong impression and believe that a 65 to 70% response rate could be achieved with paroxetine without the education and supportive psychotherapy that the placebo-treated patients in this study received. That outcome is particularly worrisome in this era of health cost containment. Thus, this study could do more harm than good unless the authors devote much more attention in their discussion to the fact that the bulk of the effect in this study was the result of good clinical management and not the medication.

   Suggested revision: Expand paragraph #2 in the Comment section to more specifically address reasons for high placebo response and include a paragraph on the limitations of the study design.

2. While the above issue is the most pressing, there are several other methodological issues with this paper including possible reasons for the high placebo response rate. The investigators apparently were permitted to include patients with conduct and oppositional defiant disorder. One of the papers cited by the authors (Hughes et al, 1990) reported that such patients have a high placebo response rate and a low response rate to imipramine.

   Suggested revision: Address this in the expanded paragraph #2 in the Comments section.

3. Another contributor to the placebo response rate is the inclusion of subjects with a 17-item Hamilton Depression Rating Scale score of 12. Many patients with a value of 12 on this scale would be considered responders in most clinical trials of antidepressants. The authors do acknowledge this point but do not address why they chose to include such patients or how many such patients entered the study or whether they were equally distributed between the three conditions or what happened to the results if these patients were excluded.

   Suggested revision: The reasons for choosing a HAMD-17 score of ≥12 as the minimum entry cut-off could be addressed in the new paragraph on study limitations. Of note: the mean (SE) baseline HAMD total scores for the 3 treatment groups are similar and are itemized in Table 2.

4. Another issue that bears on the magnitude of the drug-placebo difference is the time course of response. It is conventional in such studies to show a plot of response versus time for the various conditions to allow the reader to judge when the active treatment separated from the placebo control condition. The authors provide no information about whether

paroxetine separated from placebo only at the end of the study or at several different time points in a temporally consistent manner. This information is particularly important given the marginal drug-placebo difference. The authors should address this issue and provide at least one figure showing the time course of response.

Suggested revision: There is data available (Table 20; page 76) in the clinical report from which a figure could be constructed to show the time course of reduction in the HAMD total score.

5. The dosing of imipramine. Dosing with this drug did not employ therapeutic drug monitoring to adjust the dose to control for substantial interindividual variability in its clearance. It also involved a forced titration schedule which was slow at the beginning such that most patients would have been underdosed with this drug for the first two weeks and yet required achievement of a dose of 200 mg/day by the end of week four. This dose would be high for many patients. Yet, the authors required that patients who could tolerate such a dose had to be withdrawn from the study. This dosing schedule and requirements are such that the study was biased to find imipramine both ineffective and poorly tolerated. In contrast, patients on paroxetine were started and maintained for four weeks on its usually effective antidepressant dose based on studies in adults. It could be argued that the above is simply a reflection of the ease of optimal dosing with a serotonin selective reuptake inhibitor such as paroxetine in contrast to a tricyclic antidepressant such as imipramine. However, therapeutic drug monitoring has been used for several years in both adults and children to rationally adjust the dose of imipramine and other tricyclic antidepressants. In fact, the authors did monitor plasma levels of imipramine at weeks 4 and 8 but did not report the results.

Suggested revision: The Methods section could be revised to state that the plasma concentration findings will be reported in a separate publication. (Reviewer #2 also mentioned this issue)

6. The high dose of imipramine employed in this study likely also comprised the blind. The authors do not address this issue. However, the anticholinergic adverse effects cited in Table 5 are such that one would expect the authors should have been able to determine who was on imipramine with reasonable certainty.

Suggested revision: This could potentially be addressed in the new paragraph about study limitations in the Comments section.

7. Overencapsulation (page 9) is not an ideal way to pursue the blind of a study. Many patients will open the capsule to see what medication they are taking.

Suggested revision: Paragraph #2 (page 9) could be revised as follows: "Tablets were overencapsulated in matching Supro B locking capsules to preserve medication blinding."

8. The definition of remission and response overlaps in this manuscript (page 10).

Suggested revision: Reviewer #2 made a similar comment. This needs to be corrected (if inaccurate) or clarified in the manuscript.

9. The blood pressure parameters given on page 11 do not make sense (ie, systolic blood pressure >140 mm Hg/diastolic blood pressure <85 mm Hg). The authors should clarify.

Suggested revision: Reviewer #2 made a similar comment. The manuscript accurately reflects the data in the clinical report. This needs to be corrected (if inaccurate) or clarified in the manuscript.

Reviewer #2.

The authors describe a multi-site parallel groups designed study of paroxetine, imipramine, and placebo treatment of adolescent depression. Most indicators of efficacy showed a significant improvement after treatment with paroxetine relative to placebo. There were no significant improvements on imipramine relative to placebo. There was an overall order effect in which ratings of depression were lower on placebo at the end of treatment than at baseline. In addition, more subjects treated with imipramine dropped out due to adverse effects, most notably cardiovascular changes.

The strength of the study is that it is the first replication of the efficacy of antidepressants in treatment of adolescent depression and the first report of efficacy of paroxetine. The introduction does an excellent job of discussing the past studies of adolescent depression and in describing the limitations of all but one of those studies. The study is well-powered for demonstrating efficacy of paroxetine, but not for a weaker treatment effect, such as in the treatment with imipramine, due to placebo effect typical of antidepressant trials. The study design is standard for a clinical trial with use of well-standardized diagnostic and outcome measures.

A major weakness of the report is the implication that paroxetine is superior to imipramine on the basis of significant evidence of efficacy for paroxetine relative to placebo, but the absence of a significant difference between imipramine and placebo. If the intent is to compare paroxetine and imipramine, then a significant difference in the response between those two treatments must be demonstrated. Such an analysis appears not be have been planned. If demonstration of lack of efficacy of imipramine is intended, more analysis of the power of the study to show that effect should be provided. Considering the lack of efficacy is likely a lack of power, considering a high placebo response rate, the title should be changed to "Efficacy of Paroxetine in the Treatment of Adolescent Major Depression: A Randomized Controlled Trial." If the authors wish to continue to emphasize the lack of efficacy of imipramine, they need to demonstrate greater than 95% power (to adhere to the standard of 5% alpha level for a positive statement - given that the null hypothesis in lack of efficacy is presence of efficacy, the authors would need to design a study to show that imipramine isn't efficacious given 5% power). This is particularly important given the absence of a report of the TCA levels obtained and a relatively low administered dose. Weight range should be provided in description of the three treatment groups. Given 70 kg subject weight, a dose of 200 mg would be less than 70 kg. Also, the authors should clarify that although up to 300 mg was administered, subjects would have had steady state levels based on 250 mg dose for only 3 weeks and on 300 mg for only 2 weeks. Therefore, the comparison with paroxetine may have been designed for paroxetine to be at a more optimal dose than imipramine, further undermining confidence in an assertion of differential efficacy.

Suggested revision: 1) Be crystal clear that this study did not compare paroxetine with imipramine; 2) discuss the imipramine dose in the new "limitations of study" paragraph in the Comments section.

The study provides extremely useful tables in showing adverse events of paroxetine and imipramine in comparison to placebo. In addition to previous studies of TCAs, these data add to the overwhelming evidence of increased cardiovascular events and dropouts in treatment with tryciclic antidepressants. It would be easy to conclude that TCAs should no longer be considered first line treatments for adolescent depression and that is implied in discussion of whether subsequent trials of TCAs will be performed. However, there is a major omission from the tables. The serious adverse events should be at the top of any table of adverse events and these do not favor paroxetine. In fact, it is troubling that the authors do not note a significant increase in SAEs after paroxetine (but not IMI) relative to placebo ($p<0.05$ by Fisher's exact test). Most importantly, many have assumed that with fewer cardiovascular side effects, TCAs are safer to prescribe. However, given the high rate of primary care prescription of antidepressants and the readership of *JAMA*, it is important to emphasize that behavioral side effects in a minority of patients treated with paroxetine may be more serious than with TCAs and that they require excellent provision of psychiatric assessment and management, including access to psychiatric hospitalization. In other words, it is easier to assume quality control for ECG administration and reading than to know that all of the primary care physicians prescribing antidepressants have adequate training in monitoring of the psychiatric side effects of SSRIs and other antidepressants.

Suggested revisions: 1) This reviewer wants the serious AEs listed in the table, but these are detailed in full in the text; 2) The issue of behavioral side effects should be addressed in the Comments section (page 18; paragraph 3).

It is also easier to assume access to ECGs than weekly supportive clinical visits with experts in treatment of adolescent depression. The authors do not sufficiently highlight that the level of psychological treatment provided in this study is much more intense than that covered by almost every health insurance plan and far exceeds the usual time spent between a primary care physician and a depressed patient given continuing pressure from third party payers and ongoing discrimination against psychiatric patients and psychiatric treatment (provided by generalists or psychiatrists).

Suggested revisions: Because of the level of supportive treatment received by all groups, including the placebo group, the findings of this study may not be directly applicable to routine primary care settings. This point should be made in the Comments section.

The protocol does not exclude prior use of imipramine or paroxetine, other than recent use or an adequate trial within 6 months. This may allow inclusion of either past responders or past non-responders. The number of patients treated with IMI or paroxetine in the past should be listed.

Suggested revision: If available, the number of patients having prior treatment with study drugs will be stated in the demographic section of the Results.

It is not clear why 21 authors are given publication credit, but 9 are only acknowledged. Given the control by the sponsor of the study, apparent conduct of data analysis, and its publication, the reason for the two authors

at the sponsor's site being given authorship credit and the professionals not included should be justified to the journal upon submission. Given concern about the autonomy of the authors and sufficient input into the analysis and interpretation, the authors should state that all authors were granted full access to the full data set to verify the accuracy of the report, that all authors were in full agreement with the manuscript as submitted, or what mechanisms were provided for resolving disagreements, particularly when they involved discrepancy between views of investigators and the sponsor.

Suggested revision: A statement, such as "Funding for this study was provided by SmithKline Beecham Pharmaceuticals; each author had full access to all data and signed-off on the manuscript before it was accepted for publication.", could be added to the Study Design paragraph on page 7.

Minor points - Reviewer #2:

1. The use of the term adolescent is based only on an age of 12-18. Some of the younger boys may have been prepubertal. Adolescent is defined by post-pubertal status. The Tanner stage of all participants should be included in the description of the subjects.

Suggested revision: Tanner stage data, if available, could be added to the demographics portion of the Results section.

2. Throughout, the term effectiveness is sometimes used when efficacy is what was being tested.

Suggested revision: The word efficacy can be globally substituted for effectiveness throughout the manuscript.

3. p.5, Results - 'improvement in all treatment groups' should be reworded

Suggested revision: This revision will be made.

4. p.5, Results - if a statement is made about increased drop-out from imipramine, an analysis showing this is significant should be provided in the body of the report.

Suggested revision: This issue will be addressed by the SB statistician.

5. p.5, Conclusions - 'optimal dose' implies a single dose rather than determining the range of optimal doses across adolescents.

Suggested revision: This revision will be made.

6. p.6, para 2 - provide the median and range of previous sample sizes of TCAs in adolescent depression.

Suggested revision: Making this revision will make the sentences in this paragraph very cumbersome.

7. p.7, para 1 - 'Another study, employing a historical…' is confusing following the previous sentences and is probably best demarcated by a new paragraph or other indicator of transition.

Suggested revision: This revision will be made.

8. p.8, para 1 - The PPVT is not an intelligence test, so it should not be described as an IQ score. It should be described as PPVT standard score of at least 80. It could be further described as an indicator of an aspect of language relatively well correlated with IQ. However, many patients within the higher end of mild mental retardation will have a PPVT of at least 80.

Suggested revision: This revision will be made.

9. p.8, para 3 - change 'pervasive mental disorder' to 'pervasive developmental disorder.'

Suggested revision: This revision will be made.

10. p.9, para 2 - it appears that placebo was not administered during the screening phase, but this should be clarified and a comment should be provided later on the advantages and disadvantages of not having a placebo run-in, given comment on this by several of the authors in other publications about this topic.

Suggested revision: This revision will be made on page 9 and on page 17 (Comments section).

11. p.11, para 2 - as further evidence of not fully testing efficacy of imipramine, it is not clear why patients with TCA levels greater than 500 were dropped from the study rather than having dosage adjustments. Also, the authors should comment on whether GCP was followed, if patients were not tested for levels 1 week after dosage change or initiation of treatment with TCAs, given that if a subject had a lever of >500 ng/mL at the end of week 4, they likely had increased levels at the end of week 1 or 2. Range of TCA levels at the end of week 4 and the end of week 8 should be provided. The number of subjects excluded with levels >500 ng/mL should be provided.

Suggested revision: 1) Reasons why patients with TCA levels >500 were withdrawn and state how many; 2) The Methods section could be revised to state that the plasma concentration findings will be reported in a separate publication. (Reviewer #1 also mentioned this issue).

12. p.11, para 2 - it is likely that the authors didn't exclude normotensive adolescents, so it is assumed they meant to exclude subjects with diastolic blood pressure >85 mm Hg.

Suggested revision: Reviewer #1 made a similar comment. The manuscript accurately reflects the data in the clinical report. This needs to be corrected (if inaccurate) or clarified in the manuscript.

13. p.12, para 3 - since family history wasn't described in the methods, it is unknown what the authors mean by positive family history. Presumably, this is any relative, rather than first-degree relatives, but it should be clarified.

Suggested revision: This revision will be made.

14. p.13, para 2 - detail the cardiac adverse events leading to premature discontinuation.

Suggested revision: This revision will be made.

15. p.13, para 3 - clarify whether LOCF or completer analysis is being described throughout the results and table describing results.

Suggested revision: This is already addressed in the manuscript - in the statistical analysis section.

16. p.17, para 2 - description of 'numerically superior' is not appropriate and results should be described as superior only when significant. There is a bias in reporting paroxetine results as numerically superior but failing to emphasize this is also the case for many of the outcome measures with imipramine.

Suggested revision: This revision will be made.

17. p.18, para 1 - the authors do not address why comparison to buproprion isn't possible since it is already available rather than NE specific reuptake blockers.

Suggested revision: This revision will be made.

18. p.18, para 2 - Dose-finding was inadequate for making comment on doses administered.

Suggested revision: Disagree - this revision will not be made.

19. Table 3 - There is no mention in the text of the failure to demonstrate efficacy for the quality of life measures indicated in this table.

Suggested revision: The reviewer is mistaken; this is already in the text of the manuscript.

20. Figure 2 - there are two bar graphs, but the p values apear to only refer to one of them.

Suggested revision: The reviewer is mistaken; Figure 2 is correct.

Reviewer #3 (Statistical reviewer)

4,1 From the way the last sentence is worded, it appears as though a treatment (SSRI) is being compared to a comparison (placebo vs. a tricyclic antidepressant). This is clarified later in the manuscript, but at this juncture it is unclear.

Suggested revision: Be crystal clear that this study did not compare paroxetine with imipramine.

4,2 The wording (likewise on page 7) suggests that the combination of paroxetine and imipramine is being compared to placebo. Again, this is clarified, but only later.

4,5 How was the dose (20 mg to 40 mg) of paroxetine chosen for a specific patient?

Suggested revision: This reviewer is mistaken and is misreading the abstract.

5,1 The fifth efficacy endpoint, CGI, groups very much improved with much improved. If we can assume that very much improved is better than much improved, then combining these categories is tantamount to throwing away data which can be used to distinguish among different outcomes. This would be an inappropriate dichotomization of what is at least a trinomial endpoint. See Moses, L. E., Emerson, J. D., and Hosseini, H., 1984, "Analyzing Data from Ordered Categories," *New England Journal of Medicine* 311, 442-448. At the very least, patients could be classified as "very much improved," "much improved," or "less than much improved." Then you would use a single comprehensive analysis, such as the Smirnov two-sample test. See Berger V. W., Permutt T., and Ivanova A., 1998, "The Convex Hull Test for Ordered Categorical Data," *Biometrics* 54, 1541-1550).

Suggested revision: This revision will be addressed by the SB statistician.

5,2 Significantly greater improvement than what?

Suggested revision: This revision will be made.

6,1 What is the meaning of lifetime prevalence for an adolescent?

Suggested revision: This revision will not be made.

9,2 Was placebo administered during the screening phase? If so, then were responders to placebo excluded? If so, then this should be made explicit in the interpretation of the results. See "Run-In Periods in Randomized Trials," Pablos-Mendez et al., JAMA 1/21/98, 279, 3, 222-225 and "Threats to the Validity of Clinical Trials Employing Enrichment Strategies for Sample Selection," Leber P. D. and Davis, C. S., *Controlled Clinical Trials* 19, 178-187, 1998.

Suggested revision: Placebo was not administered during the screening phase, and this will be emphasized in the text.

9,2 What determined the length (7-14 days) of the screening phase for a patient?

Suggested revision: This revision will not be made.

12,1 It is deceptive to refer to an analysis population based on having at least one post-baseline efficacy evaluation as "intent-to-treat." The true intent-to-treat population consists of all patients randomized, analyzed as they were randomized. See Heitjan, D. F., 1999, "Causal Inference in a Clinical Trial: A Comparative Example," *Controlled Clinical Trials* 20, 309-318.

Suggested revision: This revision will be made.

12,1 A sensitivity analysis should be performed, using other imputation methods.

Suggested revision: This revision will be addressed by the SB statistician.

12,2 Were the ANOVA assumptions checked? What were the results?

Suggested revision: This revision will be addressed by the SB statistician.