This study involved an 8-week multisite randomized double-blind design with parallel arms, comparing both imipramine and paroxetine to placebo. Paroxetine was superior to placebo on some of the clinician ratings used as outcome measures. There was a high placebo response rate. Imipramine was not superior to placebo, and had a high rate of adverse effects.

This study has multiple strengths, including large sample size, randomized controlled design, and the use of standardized measures addressing multiple domains. Moreover, the study addresses an important area of clinical child psychiatry, the efficacy of antidepressant therapy in depressed youth. The results are clearly presented. Documenting that paroxetine has efficacy in adolescent depression is an important finding.

There are some issues that if addressed would greatly improve the paper. Although the authors devote considerable discussion to the high placebo response rate, this primarily serves to deflate the validity of the paroxetine results, rather than truly explore the significance of these findings. Several issues are seemingly ignored, included:
1. The implications of a high placebo response rate given that the average subject was depressed for one year prior to entry into the study.
2. The fact that parent and subject ratings did not differentiate active medication from placebo.
3. The fact that only a few of the clinician ratings differentiated paroxetine from placebo. For example, the total HAM-D scores showed very little difference, either clinically or statistically.

There are several interesting ramifications of these results that are ignored. The authors dismiss the placebo response rate as consistent with findings in adult studies, but that is not accurate. The issue of high placebo response rate in youth is a very consistent finding that has generally been attributed to sampling or severity issues. The field has maintained that the diagnosis of depression in youth is essentially the same as that in adults, yet the treatment literature is strikingly different. Do these findings potentially suggest something about either the method used to diagnose depression, or the validity of the diagnosis itself, in this population? How is it that youth with persistent major depression for a year improve at basically a 50 percent response rate in a placebo arm.
Overall this is an important study due to its large size and its design of SSRI vs. TCA vs. Placebo. However, the results do not clearly demonstrate efficacy for paroxetine. Therefore, the authors need to clearly state this.

ABSTRACT: As mentioned above, efficacy was not demonstrated for paroxetine. It should be explicitly noted that paroxetine was not found to be superior to placebo on 17 item symptoms of suicide depression on the Hamilton depression subscale. The authors might consider why these findings were unexpected in the Conclusions section.

INTRODUCTION: The placebo used in the introduction is a good one. However, the authors should cite that many of the references used are review articles and not original communications of scientific data.

METHODS: Since most of the readership may not be familiar with the supportive care management provided, a more extensive description is indicated. As this was a pharmaceutical industry-sponsored study, it is likely that there was a primary outcome measure that was identified a priori. If so, the authors should clearly state what that primary outcome measure was in the METHODS, RESULTS, and DISCUSSION sections. Since there was a large number of depression outcome measures used and because a Bonferroni correction was not employed, this is a particularly important consideration. As the number of outcome measures are described in the RESULTS section, a thorough description about the subset of how the number of outcome measures selected would vary throughout the study should be noted in order to facilitate the interpretation of these data.

The third paragraph on page 10 should read “If changes in cardiovascular parameters occurred, these designations were required”.

As before, considering the large number of outcome measures that were considered, the rationale for not using a Bonferroni correction should be described.

RESULTS: Based on the description, it may be more appropriate to note in the Adverse Events subsection that paroxetine was “generally” or “nearly” well tolerated. In addition, it should be noted how the severity of the AEs was defined or operationally. With mention of a “tolerable” versus a “moderate” adverse event! A statement is made about “side effects” on p 15. There is no mention of this in the Methods section. It is not clear why patients with a serious adverse event completed treatment and were not withdrawn from the protocol. The authors should use Standard Deviation in Tables 2 and 3, as these tables are more informative and more appropriate.

Continue on separate sheet if necessary
COMMENT: The fact that this was not a placebo-controlled study should be considered an important limitation in judging the usefulness of MDD as a whole. It is clear that this combination of a drug that is quite effective in treating depression in other studies should not be considered a substitute for this study. Therefore, the authors should not overestimate the efficacy of the treatment. The fact that there was no significant difference between the two groups is quite unusual for an industry-sponsored study. If this is the case, this study should be deemed as a methodological shortcoming.

If there was a "primary" outcome measure, the authors should clearly state what that was.

The authors state the "optimal dose range" for paroxetine should be an area of further study. They should note that there are 2 reports that describe 10 mg of paroxetine as optimal for most patients with MDD (Waxman et al, 1999; Pfeiffer et al, 1999). It is possible that 10 mg might have been the optimal starting dose for this study and should be considered in the discussion. Were there any open label data available to the authors regarding the use of paroxetine? If yes, it is important to note that the drug was indeed the appropriate starting dose when this study was designed.
Were they really persistently depressed over that entire year? In clinical settings most “depressed” youth have moods that are more labile and reactive. Is this well reflected in methods using standardized interviews? Is there possible improvement due to the way either parents or youth report depressive symptoms? A more broad based discussion of these issues that challenged existing dogmas would be very interesting and of great benefit to the field.

There are other issues that need addressed, including:

1. The rate of serious adverse events in the pharmacotherapy arm is somewhat high (11 subjects, presumably out of 93). This is included in the results, but not discussed at all.

2. Similarly, each group had a fairly high rate of not completing the 8-week trial that is somewhat glossed over.

3. Given the high placebo response rate, what algorithm should clinicians follow when treating a depressed teenager. Are SSRI’s an acceptable first line treatment if approximately one-half of youth get better with only supportive interventions.

4. Although it implied, a stronger statement could be made regarding the lack of indications for tricyclic antidepressants given the lack of efficacy and side effect profile.

5. In the discussion section, there is a statement suggesting that a traditional three arm comparative trial was not done due to the risk of exposing additional subjects to clinical research. This sounds rather self-serving, since I suspect the power issues and sample size limitations prevented this from being done, not human subjects concerns.

6. In the discussion section, there is a statement that the entry HAM-D score required was lowered to greater than, or equal to 12 "to reflect the severity of their disorder" in a pediatric population. What does this mean? Is the scale not valid in youth? Are their scores somehow different than adults?