Reviewer's report

Title: A double-blind, randomized, placebo controlled trial of cognitive-behaviour therapy, alone or co-administered with quetiapine, in refractory depression

Version: 1 Date: 26 November 2007

Reviewer: Jonathan W. Stewart

Reviewer's report:

Chaput & Magnan: A double-blind, placebo-controlled trial of . . .

These authors took on the difficult task of determining what to do next for patients who are still not well following two well-utilized antidepressants plus lithium augmentation. They demonstrate that adding quetiapine to CBT is superior to next using CBT alone. Unfortunately, they conclude that both CBT and lithium are useful treatments for doubly refractory depressed patients, but without supportive data.

I believe these authors suffer from a failure of logic, or at least a lack of understanding of statistical inference. Thus, just because a few patients improved somewhat on some rating score does not demonstrate much of anything (especially that the intervening treatment actually did anything). Rather, it is only when one group randomly assigned to double-blind treatment with X does statistically better than a comparable group assigned to Y that we can suggest anything about efficacy. They do not seem to understand that demonstration of change over time does not equal demonstration of efficacy for the treatment applied in the interim.

This sort of imprecision permeates the manuscript. These are correctable flaws, however, that do not detract from the possible contribution their study might make to a prevalent but largely unstudied problem.

Title

Might be considered misleading, as although double-blind, randomized and placebo-controlled, CBT was given to everyone, so it is not a double-blind, randomized, placebo-controlled study assessing the efficacy of CBT. Rather, it is a double-blind, randomized, placebo-controlled study assessing the efficacy of quetiapine (in the context of people also receiving CBT).

Introduction

While ‘efficacious” is one way to interpret STAR*D results, it is hardly the only one. Alternatively, perhaps STAR*D merely demonstrate that none of the treatments altered the natural fluctuation of illness differently than any of the others. Because none of the STAR*D treatments demonstrated differential efficacy relative to the others, perhaps it is more plausible to assert that none upset the natural course of illness, than that all “work” to exactly the same
degree. Again, perhaps they could state things less misleadingly in the second paragraph.

The authors’ stated primary aim was to “assess the effectiveness of CBT”. The Methods do not allow one to judge this, so it seems a peculiar primary aim. As the only thing one can judge from their methods is whether quetiapine can improve on the efficacy of CBT alone, I wonder why the effectiveness of quetiapine (or quetiapine as an add-on to CBT) is not the primary aim. Instead, this is listed as a secondary aim.

Methods
I wonder how many patients had entry HAMD = 20, and randomization = 18. I.e., how many had their scores inflated in order to shoe horn them into the study? Forced scores ought to drop once study entry constraints are removed, independent of treatment effects. Drops in post-randomization scores then become misleading in suggesting treatment efficacy. Thus, apparent improvement in patients treated with CBT alone might be partially attributable to such obligatory post-randomization drops in HAMD scores. Post-randomization changes in scores not required for randomization seem more meaningful. This seems especially important in their 31 patients whose mean HRSD was only 23, so presumably including a lot who barely met study entry criteria.

They state they excluded patients with most of the common psychiatric disorders, including personality disorders. They do not state how they determined whether their patients had personality disorders. As the more reliable methods are quite time consuming, one wonders how well personality disorders were excluded. But, also, why, as apparent personality disorders sometimes disappear once a patient is no longer depressed.

I doubt they excluded patients allergic to penicillin, yet state that patients with “nonwn drug allergies” were excluded. Granted, the offensive sentence ends with the clause “that might contraindicate . . .” But, why not clarify that it is specifically a quetiapine drug allergy that they intended to exclude.

As best I can tell, they recruited 40 patients who had previously been treated with maximal recommended doses of at least two antidepressants who were then treated openly with lithium, non-responders (defined) were then washed out of current antidepressant plus lithium and if still meeting entry criteria then randomized. So, their group really consisted on twice antidepressant refractory plus lithium augmentation refractory subjects.

They state that “maximal recommended doses” were required to consider patients to be refractory. However, they never mention what medications had been used or at what doses or whose “recommended doses” were utilized. Perhaps these were ideosyncratic doses I (and possibly others) would consider grossly inadequate.

Results
After demographics and illness history, next thing should be to document what
the prior medications were that patients were deemed to be refractory to and what the maximal doses of these had been that they had taken for at least three weeks each. Otherwise, the reader has no idea whether he or she would agree that these are “treatment refractory” patients according to their own definition.

They report screening but not randomization scores.

Table 2 is a peculiar way to report study outcome. Why not means, change scores, % whose score changed by X, or % meeting some final criterion score, such as HDRS < 8 or CGI-I of 1 or 2? None can be inferred from their presentation.

Since their methodology cannot address the efficacy of CBT, it only becomes relevant as a measure of quetiapine’s efficacy. Thus, I’d recommend their leading off with describing the population that got to randomization (as they do, but expanding on that), then jumping right into Efficacy of Quetiapine, only mentioning what occurred in the CBT only group as a contrast for the quetiapine group. Leave editorializing as to whether what occurred on CBT alone can be attributed to CBT vs. the natural course of illness to the discussion. It is a minor issue in this study, given its methodology, so why highlight it?

What does “Day 70” mean? Let’s see, 21 days on lithium plus 8 days tapering off earlier treatment = 29 days, so, perhaps Day 70 means 41 days into randomized treatment, or about 6 weeks of quetiapine. That seems about right, but it could mean about 10 weeks if Day 1 is first day of randomized meds.

No idea what the significance of HRSD # 14 is, as someone with a score of 14 is still has considerable depressive symptoms and I do not believe a clinician would consider to be doing well. It is ideosyncratic without any mention as to why it was chosen, why more conventional cut-offs or definitions were not used, or what it might mean clinically, other than it “feels” like a drop of at least 14 from randomization requirement must be an improvement.

They state that quetiapine was well-tolerated. If so, I wonder why the mean dose was so low and no one reached the allowed maximum in the face of most patients not remitting. They do not report what the pills per day were for those receiving placebo. If the range in number of pills equals that of quetiapine-treated patients, one wonders about the vigor of the treaters. If most reached max or at least near max, one wonders about the statement that quetiapine was well-tolerated.

I wonder how “CBT significantly . . . resulted in clinical improvement” (bottom p. 11) jibes with “no significant improvement was observed in the CBT+placebo group.”

Would a Fisher exact p be more appropriate than a #2? I believe they used an unadjusted #2, as I get a “p” of .051 with an adjusted #2. Probably they should let the reader know whether they used an adjusted or unadjusted #2, if they choose to leave it as that test.
I suppose it useful in some way to report the change in scores over time for each group. However, as one cannot sort out treatment effects from natural course of illness over time from artifact (i.e., obligatory drop in initially inflated entry scores), why start off with their uninterpretable analyses instead of those that can be interpreted?

Discussion

The above leads directly into their initial sentence of their Discussion. Their results do nothing of the kind! I fail to see how uncontrolled treatment can say anything about “viability” unless one means by that not getting in the way of natural course; but, even there, they cannot with any confidence assert that natural course would not have done better than either lithium or CBT.

They refer to a summary article that refers to a book chapter re “placebo-expectancy”. Best to use whatever the original documentation that this is the case than someone merely mentioning it without any documentation. Similarly for a subsequent use of citation 31.

They state that all 4 measures were significant for quetiapine while “only 2 of 4” were for CBT+placebo. Although they do not make the mistake of stating that this demonstrates quetiapine’s advantage, this seems to be the implication, else why would they state it as they do? If they want to state this counting of significances, they should alert the reader to the illogic of using the difference in numbers found to be as significant as an additional indication of quetiapine’s increased efficacy.

They mention various possibilities as to why quetiapine was found to be superior to placebo. But, items and subscales on their measures at least partially address these issues. Just as an example, there are three sleep items on the HRSD. They could compare the sum of the nonsleep items as a first look at whether quetiapine was “only” effective because it helped sleep. They could look at only item 1 or only the 7 “core” items of depression to give a suggestion of whether quetiapine helps depression vs. sleep vs. anxiety vs. cognitive problems. There is also an accepted anxiety factor to the HRSD they could use as a way to look at its anxiety effects.

Conclusion

Again, the first sentence seems backwards! Their results suggest the efficacy of quetiapine and cannot be used to address the efficacy of CBT. At least, I see no reason to point to anything in their data to go so far as to “suggest” to me that CBT “is clinically pertinent for the treatment of TRD.” Rather, their data suggest the pertinence of quetiapine, at least if given in the context of CBT, with the caveat that the relevance of the latter is not addressed in any way.

Figure

There was a reference in the text to a figure which appeared to be missing from the manuscript I received.

Tables
I’d like to see a better documentation of prior treatment, specifically the range of mg/d and time on maximal mg/d for each of the treatments patients were supposedly refractory to. Also, how they documented that the patients were refractory prior to coming to their clinic and receiving lithium augmentation. I have had plenty of patients who convince me that X did not help, even temporarily, but I have both chart and scale score documentation (including self-reports!) of complete sustained remission on X. Memory is too tied to mood to be reliable.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

While I do not recall having received any form of support in any form from Astra-Zeneca, I have from a variety of other companies, any of which might consider Astra-Zeneca's fortunes to impact their own. In addition, conceivably my various organizations (New York State [my employer], Columbia University [my academic affiliation] and the Research Foundation for Mental Hygiene [which processes my grants]) could potentially benefit or be hurt by the publication of this paper. Should the editors like a full disclosure of the companies that have ever given me money or studies, I will be happy to supply a list.

I hold no stocks or other types of shares in any company that could conceivably be helped or hurt by the publication of this manuscript.

I neither hold nor am I applying for any patents.

I have no idea what patents any organization I have received money from holds or may be applying for.

I know of no other financial competing interests.

I have a non-financial competing interest to the extent that this is my field of research. I do not see how the success or failure of this manuscript either helps or hurts me, but I assume a clever lawyer could.