Reviewer's report

**Title:** Paroxetine in the treatment of mildly depressed women with type 2 diabetes: A single-blind randomised placebo controlled trial

**Authors:**

Maria Paile (maria.paile@helsinki.fi)
Prof Johan G Eriksson (johan.eriksson@ktl.fi)
Kristian Wahlbeck (kristian.wahlbeck@stakes.fi)

**Version:** 1  **Date:** 27 Jan 2003

**Reviewer:** John M Hughes

**Level of interest:** A paper whose findings are important to those with closely related research interests

**Advice on publication:** Unable to decide on acceptance or rejection until the authors have responded to the compulsory revisions

The trial reports the effect of paroxetine compared with placebo in reducing HbA1c and improvements in mental state in sample of mildly depressed, obese women.

**Compulsory Revisions**

1. The correct analysis must be performed.

The analyses performed in this paper are incorrect. The analysis required is to test if the change from baseline between treatment groups is clinically significant and not, as the authors have done, to test for a clinically significant change within each group [1]. Thus the main conclusion that GHbA1c 'did decrease significantly’ in the paroxetine group has not shown that the decrease was clinically significant compared to placebo. This analysis could be performed with an ANCOVA model with the baseline measurement as the co-variate giving the model:

\[
\text{Final measurement} = \text{baseline measurement} + \text{treatment}
\]

This model is preferable to the two-sample t-test on the change from baseline values or the non-parametric equivalent if the appropriate assumptions are not met.

The authors state that they have employed the non-parametric Mann-Whitney U-test and the Wilcoxon signed ranks tests for comparison of means between and within groups respectively. However the statistic reported in the Results section appears to be the Z statistic from the standard Normal distribution. It is not clear how the authors calculated this statistic. The Z-statistic is a parametric analysis that implies that the distribution of these data did not show serious departures from normality. Which raises the question why were non-parametric analyses used? But in any case the Z-statistic is inappropriate because of the small sizes of the groups.

2. What is the clinically important difference?
The authors contradict themselves over what is a clinically important decrease in GHbA1c. Their original outcome for GHbA1c was clinically important if there was a 10% decrease from baseline. However in the Discussion section they accept that although the decrease in the paroxetine group only reached a percentage change of approximately 6%, this was equivalent to an average decrease of 0.44% in absolute units now considered clinically important. If 0.44% is a clinically important difference in GHbA1c then the analysis should have been performed on the absolute values with 0.44% as the clinically important difference between the groups.

A sample size calculation based on this difference between the groups shows that approximately 100 patients are required in each arm to detect a difference of 0.44% assuming a standard deviation of 1% in absolute units. This study could detect a difference of approximately 1.5% GHbA1c assuming a significance level of 5% and a power of 80%. Hopefully a sample size calculation will be performed before the authors conduct their longer double-blind placebo controlled study.

3. Tables should report differences between treatment groups.

Tables 2 and 3 report change, standard deviation and p-values of changes within groups. These tables should report between group differences and confidence intervals rather than p-values once the correct analyses have been performed. With so few patients in the groups the confidence intervals will be very wide but this will be more illuminating than simply reporting the p-values.

The above comments have used GHbA1c as the example but they also apply to the other parameters reported by the authors. The authors have performed many significance tests and this can produce spurious significant results. However if their analyses are confined to between group differences the number of tests needed will be halved.

4. Were the investigators blind to the treatment?

The authors need to clarify why the study was single blind. In a single blind study the patient doesn't know which treatment they are receiving. In a double blind study neither the patient nor the clinical assessor know which treatment the patient is receiving. How was it that the assessors were not blinded to the treatment? In any case, for the metabolic parameters, did those who performed the laboratory analyses know the treatment each patient received?

5. Greater detail is needed for adverse events.

The statement on adverse reactions is completely uninformative because it fails to specify what these were or whether they were related to the study drug. Furthermore it only gives the number of patients who suffered adverse reactions and not the number of adverse reactions reported, although this might be the same.

Discretionary Revision


The authors have tested baseline characteristics between the two treatment groups. As the usual significance level is 5% this test will produce a significant result in one in twenty tests. Nothing is achieved by performing it.

Competing interests:

None declared.