METHODS APPRAISAL
IN PSYCHIATRY

MANUSCRIPT AND APPRAISAL DETAILS

APPRAISAL: 2012/08/jad-adult-adhd
TITLE: Duloxetine in Adults With ADHD: A Randomized, Placebo-Controlled Pilot Study
AUTHORS: Mathieu Bilodeau, Tarek Simon, Miriam H. Beauchamp, Paul Lespérance, Simon Dubreucq, Jean-Pierre Dorée and Smadar Valérie Tourjman
SPONSOR: Psychiatry Department of the Centre Hospitalier de l’Université de Montréal and personal research funds
JOURNAL: Journal of Attention Disorders
RANKING: # 1 most downloaded on MedLinx (psychiatry) in June 2012

CONSORT CHECKLIST

C2.a BACKGROUND - Scientific background and explanation of rationale
Not a matter for statistical judgement.

C2.b OBJECTIVES - Specific objectives or hypotheses
✔ The overarching aim of the trial is clearly spelt out in the Introduction; “The aim of the present pilot study was to investigate the effect of duloxetine in adults with ADHD ...”

☒ This broad aim is further narrowed down to a primary objective which includes 2 outcome measures. From the Discussion section: “The primary objective was to test the hypothesis that 60 mg of duloxetine daily is superior to placebo in the treatment of adult ADHD, as measured by changes on the CAARS-Inv:SV and CGI scales.” This is a concern; “test the hypothesis” and “superiority” is the language of the phase III & IV clinical trial. Possibly the most common error observed in pilot studies is misplaced emphasis on statistical significance, instead of feasibility (the pop-gun fallacy).

Pilot or feasibility studies by definition do not intend to demonstrate superiority; their more modest aims are to explore relevant clinical issues & logistics and to gather useful information and insights to aid the design of a subsequent larger, fully powered, confirmatory study.

☒ Suggestion: The primary objective of this pilot study should be to measure observed treatment effect and its variation, not to engage in futile hypothesis testing. Observed treatment effects and the amount of variation should be reported, not p-values and claims of “significance”.

☒ In our opinion the secondary objectives should be focussed at least partly on the feasibility of conducting an antidepressant trial for ADHD.

☒ Suggestion: Create secondary objectives to measure the willingness of adult ADHD patients to take enter an antidepressant trial, the refusal rate, the observed rate of accrual and the observed 6 week treatment adherence - this is useful information regarding the feasibility of conducting a larger trial.

Appraisals by statisticians in plain English for clinicians
By choosing to use 2 outcome measures in effect 2 specific hypotheses are now tested as the primary objective – in other words co-primary objectives. Furthermore in the Results section both CGI-S and CGI-I scales along with both subscales and total score of the CAARS-Inv:SV are reported; in total 5 outcome measures. This results in confusion as to which outcome measure is the primary outcome measure. This also makes it hard to determine if the pilot was a success or not.

Suggestion: Choose a single outcome variable for the primary objective. Relegate the others to supportive secondary analyses. If there is a compelling reason for co-primary objectives this should be explained in the Methods section. Rank the secondary objectives in order of importance (report the secondary objectives in this order).

Some of the secondary objectives specified in the clinical trials registry go unreported in this manuscript.

The manuscript would benefit from a criterion for a successful pilot.

C3.a TRIAL DESIGN – Description of trial design (such as parallel, factorial) including allocation ratio

✓ All relevant details are clearly spelt out in the Methods / Study Design section.

C3.b CHANGES TO TRIAL DESIGN - Important changes to methods after trial commencement (such as eligibility criteria), with reasons

NA

C4.a PARTICIPANTS - Eligibility criteria for participants

✓ All relevant details are clearly spelt out in the Methods / Participants section.

C4.b STUDY SETTINGS - Settings and locations where the data were collected

✓ Single site noted in the Methods / Study Design section.

C5 INTERVENTIONS - The interventions for each group with sufficient details to allow replication, including how and when they were actually administered

✓ Details are clearly spelt out in the Methods / Dosing subsection.

C6.a OUTCOMES - Completely defined pre-specified primary & secondary outcome measures, including how and when they were assessed

✓ Primary and secondary outcomes are outlined in the Methods / Assessments subsection.

✓ Full details of the primary model are not given in the Methods section “… were analysed with repeated measures ANOVA.” is too vague.

Suggestion: “A repeated measures model was fit to the longitudinal outcome data… (fully describe model here…) The treatment group-by-time interaction term (this describes the time course) in the model was scrutinized for evidence of a treatment effect.”

✓ The secondary outcome measure in the clinicaltrials.gov registry (CGI) is now elevated to a co-primary outcome measure as stated in the manuscript’s Method / Assessments section.
The co-primary efficacy measures stated in the Method / Assessments section are not reported in the Results section in the appropriate order: CGI-S (a stated secondary outcome measure in the registry) is the top line result while CAARS-Inv:SV is reported eventually after its two subscores. As well as being demoted to the 5th result the primary outcome measure is not plotted.

Two secondary outcome measures (HARS-14 and HDRS-17) were not reported in the Results section though they are mentioned in the Discussion. Other secondary outcomes such as quality of life (quality of life is an unusual outcome measure for a pilot study) and cognitive & executive function go unreported in the manuscript.

From a biostatistician’s perspective this situation is undesirable. It is highly unusual to re-order a trial’s objectives at any stage in during the research effort as it leaves the manuscript open to criticism and speculation they may have been re-ordered after the final results were known.

Suggestion: Even though it is a pilot study, scientific discipline must be exercised. We feel the analysis should stick to a single primary outcome measure specified in its protocol and that the manuscript should report all specified secondary outcome measures in the order they were listed on clinicaltrials.gov (admittedly journal restrictions can make this difficult).

C6.b CHANGES TO OUTCOMES - Any changes to trial outcomes after the trial commenced, with reasons

It would have been useful to explain the discrepancy in the choice and order of outcome variables between the manuscript and the clinicaltrials.gov registry.

C7.a SAMPLE SIZE - How sample size was determined

There is no formal sample size justification. We feel a sample size justification can and should still be done for a pilot study based on the precision (not power) that could be expected from the given sample size.

The Discussion section notes “This is most likely the result of lack of power to detect group differences.” The focus on power and detection of group differences in a pilot study is completely misguided.

C7.b INTERIM ANALYSES AND STOPPING GUIDELINES - When applicable, explanation of any interim analyses and stopping guidelines

NA

C8.a RANDOMISATION: SEQUENCE GENERATION - Method used to generate the random allocation sequence

No information is supplied regarding the specific details of randomisation.

Suggestion: refer http://www.consort-statement.org/consort-statement/3-12---methods/item8a_randomisation-sequence-generation/

C8.b RANDOMISATION: TYPE - Details of any restriction (such as blocking and block size)

No information is supplied regarding the specific details of randomisation.

Suggestion: refer http://www.consort-statement.org/consort-statement/3-12---methods/item8b_randomisation-type/
C9 RANDOMISATION: ALLOCATION CONCEALMENT MECHANISM - Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned.

⚠️ No information is supplied regarding the specific details of randomisation.


C10 RANDOMISATION: IMPLEMENTATION - Who generated the allocation sequence, who enrolled participants, and who assigned participants to interventions.

⚠️ No information is supplied regarding the specific details of randomisation.

⚠️ Suggestion: refer to [http://www.consort-statement.org/consort-statement/3-12---methods/item10_randomisation-implementation/](http://www.consort-statement.org/consort-statement/3-12---methods/item10_randomisation-implementation/)

C11a BLINDING - If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how.

⚠️ No information is supplied regarding the specific details of blinding. Vague use of the term “double-blinded” in the text should be discouraged.

⚠️ Suggestion: Spell out precisely who is blinded – patients, raters, analyst, etc refer to [http://www.consort-statement.org/consort-statement/3-12---methods/item11a_blinding/](http://www.consort-statement.org/consort-statement/3-12---methods/item11a_blinding/)

C11b SIMILARITY OF INTERVENTIONS - If relevant, description of the similarity of interventions.

⚠️ No information is supplied regarding the specific details of blinding. We assume the placebo tablet was of similar appearance to the active treatment.


C12a STATISTICAL METHODS - Statistical methods used to compare groups for primary and secondary outcomes.

⚠️ The chosen level of significance 0.05 is typically used in phase III (confirmatory) trials seems particularly stringent for a pilot (exploratory) trial (we’d be happy with 0.1)

⚠️ The text states a repeated measures ANOVA model was used however specific details of this model are missing from the Methods section (some details can be elucidated from the Results section) important details like whether adjustment was made for outcome score at baseline is unclear. This is a serious omission, as the model forms the backbone of this longitudinal analysis. Its precise form needs to be clearly specified to the reader.

⚠️ Suggestion: In our opinion reporting the mean change in CAARS-INV:SV score at 6 weeks along with its 95% CI would be the most appropriate efficacy variable to report for this trial. This information could guide investigators in designing a larger confirmatory study.

⚠️ Conversion to percentage change of the 6 week change in CAARS-S:L score is misguided. In our opinion this variable should be analysed in its raw form (with an ANCOVA model). It’s not clear why this analysis...
decision was made. Consider the result reported in Table 2: “Problems with Self Concept” for the placebo group +34.8% (130.11%). A standard deviation this large (130%) is impossible since the maximum change can be +/- 100%, and percentage change data is often not Normally distributed making statistics like the standard deviation inappropriate.

Suggestion: Analyse in the raw form with model:
Change score = baseline score + treatment + (other relevant variables)

C12.b ADDITIONAL ANALYSES - Methods for additional analyses, such as subgroup analyses and adjusted analyses
NA

C13.a PARTICIPANT FLOW - For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome
✓ No flow diagram is supplied, however owing to the trial’s small size, simplicity and low dropout. Participant flow is adequately described in the Results / Demographics and Baseline Values section.

✓ A futile statistical comparison was made between characteristics of completers (n=24) and those who dropped out (n=6), “no significant difference” was reported – this is the fallacy of futile significance testing.

C13.b LOSSES AND EXCLUSIONS - For each group, losses and exclusions after randomisation, together with reasons
✓ Well described in the Results / Demographics and Baseline Values section.

C14.a RECRUITMENT - Dates defining the periods of recruitment and follow-up
✓ Clearly defined in the Methods / Study Design section.

C14.b REASON FOR STOPPED TRIAL - Why the trial ended or was stopped
NA

C15 BASELINE DATA - A table showing baseline demographic and clinical characteristics for each group
✓ Baseline data are reported in Table 1 and were judged to be similar.

C16 NUMBERS ANALYZED - For each group, the number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
✓ The completers analysis approach taken is clearly stated in the Results section.

C17.a OUTCOMES AND ESTIMATION - For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)
✓ Regrettably the text emphasises the absolute group means not the difference in group means (a.k.a. the treatment effect).

Appraisals by statisticians in plain English for clinicians
Emphasis was placed on the treatment group-by-time interaction from a longitudinal model. In our opinion the difference in group means (and 95% CI) at Week 6 would be more appropriate as this information can be used to estimate the required sample size for a subsequent confirmatory study should the results be promising.

C17.b BINARY OUTCOMES - For binary outcomes, presentation of both absolute and relative effect sizes is recommended.

NA

C18 ANCILLARY ANALYSES - Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory.

✔ Both CAARS-Inv:SV subscale results are reported.

C19 HARMS - All important harms or unintended effects in each group.

✔ A Tolerability Findings section accounts for all observed safety findings.

✔ The top line safety result reports “There were no significant differences in vital signs between placebo and treatment groups.” This is a very unsafe way to conduct safety analysis. We call this the RCT safety fallacy. This hints at a fundamental lack of understanding as to the nature and limits of statistical testing.

✘ Suggestion: Do not use significance testing on safety data, a sensible, clinically focussed appraisal of the safety data is required – not a futile statistical comparison.

C20 LIMITATIONS - Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses.

✔ No specific limitations section.

C21 GENERALIZABILITY - Generalisability (external validity, applicability) of the trial findings.

Not a matter for statistical judgement.

C22 INTERPRETATION - Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence.

✔ The chosen analysis strategy of significance testing (in spite of a small sample size) has the effect of cornering the interpretation into strictly dichotomous conclusions (“significant” & “not significant”). This is not the information required from a pilot study. In an attempt to escape this self-imposed predicament the phrase “trend towards” is used in the Discussion and throughout the Results section when the p-value is slightly above the defined significance level. This language should be avoided in the reporting of any trial “... a trend toward overall group differences (p=0.082)” could just as easily be a trend away from group differences. This language is a common form of pleading that is specific to underpowered or poorly conducted studies.

✘ Suggestion: A simple solution is to simply report the confidence interval instead of the p-value, comment on the width of the interval and possibly the clinical relevance of the upper and lower confidence limits.
In the Discussion section the statement “There was no group difference on the CAARS-Inv:SV…” when in fact there was a 7.5 point difference ((33.44 – 25.67) – (31.6-31.33) = 7.5). This highlights the poverty of significance testing to qualitatively describe the observed treatment effects in pilot studies.

Suggestion: “There was a 7.5 (insert 95% CI here) point difference observed favouring the Duloxetine group…”

C23 REGISTRATION - Registration number and name of trial registry

- Registered on Clinicaltrials.gov and clearly stated in the Methods / Study Design section
  http://clinicaltrials.gov/ct2/show/NCT00940693

- From the clinical trials registry it is clearly stated that the single primary outcome measure is CAARS-O:SV scale (while the CGI-S is the top line secondary outcome measure). However the Results section leads with the CGI findings (followed by the CAARS-Inv:SV results).

C24 PROTOCOL - Where the full trial protocol can be accessed, if available

- No details supplied regarding the study protocol

Suggestion: refer http://www.consort-statement.org/consort-statement/23-25---other-information0/item24_protocol/

C25 FUNDING – Sources of funding and other support (such as supply of drugs), role of funders

- Clearly spelt out in the Study Design section.

GRAPHICS

- The primary outcome variable CAARS-INV:SL is not plotted.

- The x axis on the plots are mislabelled, we presume it should be “visits” not “time”. The plots should have “time (weeks)” on the x-axis.

- The y axis is incorrectly labelled in Figure 2 as “CGI-Severity” rather than “CGI-Improvement”.

COST

$25 US