Author's response to reviews

Title: Utility of the sore throat pain model in the multiple-dose assessment of an acute analgesic: a double-blind, randomized, placebo-controlled study of flurbiprofen 8.75 mg lozenges

Authors:

Bernard Schachtel (bschachtel@srcresearch.net)
Sue Aspley (Sue.Aspley@ReckittBenckiser.com)
Adrian Shephard (Adrian.Shephard@ReckittBenckiser.com)
Timothy Shea (tim.shea@reckittbenckiser.com)
Gary Smith (Gary.Smith2@ReckittBenckiser.com)
Emily Schachtel (eschachtel@srcresearch.net)

Version: 3
Date: 30 January 2014

Author's response to reviews: see over
To the Editorial Board

Re: Manuscript resubmission to Trials: ‘Utility of the sore throat pain model in the multiple-dose assessment of an acute analgesic: a double-blind, randomized, placebo-controlled study of flurbiprofen 8.75 mg lozenges’ (manuscript ID: 1968002794108618)

30 January, 2014

Dear Sir/Madam,

On behalf of my co-authors, thank you very much for considering this manuscript for publication in your journal – we are also grateful to the peer reviewers for their valuable feedback.

Please find uploaded the revised manuscript incorporating the reviewers’ comments (clean and tracked changes versions are uploaded). Please also see below for our responses to each comment. I hope you find the revised manuscript acceptable for publication in Trials.

I look forward to hearing from you in due course.

Yours faithfully,

Bernie Schachtel

<table>
<thead>
<tr>
<th>Reviewer comment</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reviewer 1</strong></td>
<td></td>
</tr>
<tr>
<td>The study is well designed and performed. The methods deem to be appropriate, yet not always well described. The discussion is correct and the conclusions are consistent with the results.</td>
<td>Thank you for your feedback. We have now revisited the CONSORT checklist and made further amends throughout the manuscript, particularly to the Results section (e.g. addition of 95% CIs).</td>
</tr>
</tbody>
</table>
However, the results are not always reported clearly and according to the relevant standards, such as CONSORT Statement, and this may complicate their interpretation. Provided the results reporting is amended and the outstanding questions are clarified, the publication of the manuscript can be recommended.

**Reviewer 2**

1) The authors use a test procedure to describe and compare investigational treatments under methodological restrictions that improve homogeneity and that reduce the impact of possibly confounding influences. They call this approach a “model”. I take issue with the use of such denotation. Indeed, a model is a cognitive construct that relates an experimental outcome to the underlying various effect factors (their weight and variability) and their interaction (e.g. PK-model, PKPD-model, ANOVA-model, etc.). The test procedure or routine used by the investigators does not meet this concept. I think such deflational use of the “model”-concept should be avoided (« calling my twenty years old Ford Fiesta “Mercedes” does not make it run faster ...! »).

With due respect for Reviewer 2’s comments, this “test procedure” has been termed an “acute pain model” by several authorities (in the US and Europe), pain societies, and, in fact, by Reviewer 4. The use of sore throat pain as a ‘model’ for the evaluation of analgesics has been described in scores of publications since its original publication (Schachtel BP, Fillingim JM, Beiter DJ, Lane AC, Schwartz LA. Rating scales for analgesics in sore throat. Clin Pharmacol Ther 1984;36:151–6, which received a lead article position by the Editor of CPT because he recognized it as a new pain model). Since then, the sore throat pain model has been recognized as one of two standard “acute pain models” for general pain (the other being dental pain). We leave the decision about using the term “model” to the Editor of the journal.

2) We advise the authors to cross-check their paper with the CONSORT-Guidelines; there already now is good compliance, but there

We have revisited the CONSORT checklist and made further amends throughout the manuscript, particularly to the Results section.
might be some further improvement (e.g. addition of 95% CIs).

3) Accordingly, I also recommend the treatment effects (estimated differences between active treatment and placebo) to be presented with their respective 95% confidence interval and not just with their p-value.

| 3) Accordingly, I also recommend the treatment effects (estimated differences between active treatment and placebo) to be presented with their respective 95% confidence interval and not just with their p-value. | As above, we confirm that all the treatment effects have been updated to include their 95% CIs. |

4) In order to illustrate how well the active treatment worked, I recommend plotting a Cumulative Proportion of Responders Analysis Graph (as introduced by Farrar et al., 2006 and Farrar et al, 2010) to present the data on pain relief and to calculate the absolute risk difference (ARD) and related number needed to treat (NNT) for selected cut-off levels of pain relief. This might help understanding whether and to which extent the observed effects – although statistically significant – are indeed also therapeutically relevant.

| 4) In order to illustrate how well the active treatment worked, I recommend plotting a Cumulative Proportion of Responders Analysis Graph (as introduced by Farrar et al., 2006 and Farrar et al, 2010) to present the data on pain relief and to calculate the absolute risk difference (ARD) and related number needed to treat (NNT) for selected cut-off levels of pain relief. This might help understanding whether and to which extent the observed effects – although statistically significant – are indeed also therapeutically relevant. | Thank you for this suggested addition. We have prepared the recommended graph as a new figure (Figure 5) and have discussed the ARD and NNT in the Results section. |

5) Presently, the authors conclude that the effects were clinically significant by demonstrating a statistically significantly higher proportion of patients reaching 20% pain reduction, a level which is referred to as “least minimal improvement” based on Cepeda’s paper (2003). This might be a misunderstanding of Cepeda’s paper: Cepeda et al. compared agreement between a 0–10 numeric rating scale (NRS) of pain intensity and a 5-point Likert scale from ‘no improvement’ to ‘complete pain relief’ scored every 10 minutes in postoperative patients.

| 5) Presently, the authors conclude that the effects were clinically significant by demonstrating a statistically significantly higher proportion of patients reaching 20% pain reduction, a level which is referred to as “least minimal improvement” based on Cepeda’s paper (2003). This might be a misunderstanding of Cepeda’s paper: Cepeda et al. compared agreement between a 0–10 numeric rating scale (NRS) of pain intensity and a 5-point Likert scale from ‘no improvement’ to ‘complete pain relief’ scored every 10 minutes in postoperative patients. | We appreciate this comment and have removed reference of a 20% pain reduction as a ‘minimal improvement’ or a ‘clinically significant’ effect. |
treated with opioids. For patients with moderate baseline pain, a decrease of 1.3 NRS-units (20% reduction) corresponded to ‘minimal’ improvement on the Likert scale. This means nothing more, but also nothing less than that a NRS-reduction of at least 20% is required in order for the patients to experience this as the lowest achievable improvement on the Likert scale. This is different from what might be considered a clinically beneficial extent of pain relief.

| 6) The authors based their sample size estimates on an anticipated mean difference between the treatments of 20% for the primary criterion; actually, differences are reported that were much larger – was the study overpowered? | The 20% mean difference between treatment groups is the standard minimal difference commonly used to estimate sample size. In fact, in the clinical trial a 59% greater mean pain reduction was observed over 24 hours (i.e., SPID24), indicating that the active treatment was more effective than the minimal criterion of efficacy that was assumed before the trial. (The study would justifiably be called overpowered if the differences were smaller than the 20% minimal criterion, yet “statistical significance” was declared; here we observed a much larger effect.) Additionally, the sample size of ca. 100 patients in each treatment group provided an opportunity to observe the safety of multiple doses of the test medication compared with placebo in an ample number of patients. Finally, given the natural improvement of sore throat with time, we had hoped that more patients would have persistently high pain levels so we could detect a difference between active and placebo over 7 days (in fact most patients’ pain declined after 24 hours in the trial, representing 5–7 |
Bernard P. Schachtel, MD  
4300 So. US Highway One, Suite 203  
Jupiter, Florida 33477  
e-mail address: bschachtel@SRCresearch.net

| 7) The authors conclude that there was a distinct pharmacodynamic effect profile after the first lozenge with the onset of significant differentiation at the first time of evaluation (1 hour after dosing), peak effects at 1-3 hours, and duration lasting for up to 4 hours. In order to be considered a “pharmacodynamic profile”, evaluations should have started earlier after dosing, should have been more frequent and the in-clinic stay ought to have been longer (at least extending over the observation window that is considered pertinent to such profiling). Presently, with the late start of the evaluation, relevant effects occurring earlier might have been missed; with the in-clinic stay, there is uncertainty about the use of rescue medication during most of the pharmacodynamic observation period). | We agree that pharmacodynamics effects should include early effects and have removed the term ‘pharmacodynamic profile’, as recommended. |

| 8) The authors rightly emphasize that there were no statistically significant differences between the treatment groups for the cumulative effects over Days D2-D7. They suggest restricting recruitment to patients with more recent onset of symptoms in order to account for the natural course of disease (i.e. spontaneous regression of signs and symptoms related with the primary condition). I wonder to which purpose such change | As in other acute pain trials, the comparisons made over the first 24 hours represent how effective the test medication is in patients in general with the specific painful acute condition. We agree with the Reviewer that symptoms are usually worse in the initial few days of this acute illness: in our study, most patients with painful pharyngitis showed improvement beyond this initial time period |
should be made in the procedure. Medications like the present are typically prn-medications (pro re nata – as needed); as expected, they work quite well at the start; but, why suggesting that there might be benefit in using them for more than 1-2 days? (and, because the number of patients with relatively severe sore throat pain was thus reduced, we were unable to demonstrate activity later in the study). Medications such as flurbiprofen lozenge are used for sore throat symptom relief on a pro re nata basis at any time during the acute sore throat episode. Acute sore throat typically lasts for 4–7 days, so it is possible that some patients may require symptom relief for more than 1–2 days.

Reviewer 3

Comment #1
Discretionary Revision –Should the ANOVA model used for efficacy be shown as an Appendix? Please find the ANOVA model outputs for the primary endpoint (SPID24) at the end of this document.

Comment #2
Major Compulsory Revision –page 9 –could you please explain how the following statistical assumption with regards to the ‘relevant baseline included as a covariate’ was assessed: “If a patient used rescue medication for pain, all subsequent STPIS, DSS, and SwoTS scores in the 24-hour interval were assigned the baseline value”. What baseline was then assigned to subjects with no rescue medication?

If the patient used rescue medication, then the subsequent post-baseline values were imputed as the baseline value (baseline observation carried forward). If they did not rescue, then the relevant patient-reported outcome for each assessment time-point was used, and no imputation was needed. This has been clarified in the Methods section (page 9, lines 11–12, 17).

Comment #3
Minor Essential Revisions –page 9 –could you please explain rationale for imputation of missing data using the nominal time since first dose?

The primary endpoint was the SPID (summed pain intensity difference) over 24 hours, which is an approximation of the area under the change from baseline curve (AUC). When there are missing data, this method can lead to biased estimates of the area. Therefore,
linear interpolation was used to give a more reliable approximation of the AUC from the non-missing data. This has been clarified in the Methods section (page 9, lines 17–21).

**Comment #4**

Minor Essential Revisions—page 9—consider rewording the following, it is difficult to understand that logistic regression applies to the question ‘20% change reduction (YES/NO)’: “The percentage of patients who reported at least minimal improvement (at least a 20% reduction) [25] in symptom severity from the first dose on the STPIS, DSS, or SwoTS was assessed and significance was calculated using logistic regression with fixed treatment and site effects.

Logistic regression can be applied to binary (yes/no) data and allow comparisons between treatment after adjusting for the site effect. The odds of achieving a 20% reduction for a flurbiprofen patient are compared to the odds for a placebo patient after adjusting for the site effect. This has been clarified in the Methods section (page 10, lines 6–10).

**Comment #5**

Minor Essential Revisions—page 11—where are results from ‘sore throat relief’ shown? I can’t find them on any tables or figures. Consider adding them as an Appendix too.

The sore throat relief results are presented in the Results text on page 12, lines 13–16.

**Reviewer 4**

1) This report is a clearly conceptualized and articulate account of the extension of a well-established assay methodology previously employed to assess the efficacy of single doses, now extended to assess the efficacy of multiple doses over 24 hours. The first author has at least 25 years of experience with the acute sore throat model and has published results based upon this model in multiple peer-reviewed journals. The above said, my ability to see how the present manuscript differed from an earlier one, that according to the cover letter had

Thank you for your comments. Indeed, previous studies have demonstrated the efficacy of a single dose of flurbiprofen lozenge. This present study differs as the primary outcome measure was efficacy of multiple doses over 24 hours; therefore, this study focuses on the first 24 hours of treatment. This has been further clarified in the Introduction (page 5, lines 17–18).
been modified in accordance with a reviewer’s requests, was prevented by the absence of information in the cover letter concerning specific changes, nor any indication of same in the manuscript text itself.

2) As I read through the manuscript initially, my principal concerns as to analgesia for the placebo group, and detection of treatable bacterial tonsillo-pharyngitis, were only partially addressed, The free access to acetaminophen sufficed with respect to the former concern. As to the latter, the authors should insert an extra sentence indicating how patients whose throat cultures were positive for beta hemolytic streptococcus were directed to antibiotic therapy, or if not then why not. They should also indicate explicitly that (as I believe to be the case) such patients with positive throat cultures were maintained in the intent-to-treat group and the group results include these.

Patients whose throat cultures were positive for beta hemolytic streptococcus were given antibiotic therapy. Patients with group C streptococcal infection who had not improved by the time of the culture report (usually by 48 hours) were also treated with antibiotics. All eligible patients, including those with positive throat cultures were maintained in the study. This information has been added to the Methods (page 8, lines 7–11).

3) On page 7, first sentence under Assessment, the sudden introduction of the Practitioner’s Assessment of Inflammation scale -- abbreviated as "PAIN" is confusing, particularly since the present report focuses upon pain and its treatment, Throughout the entire manuscript, the Practitioner’s Assessment of Inflammation should be abbreviated in a way to minimize confusion with "pain". This referee would suggest the abbreviation "PrAoI".

We agree, the term PAIN may cause confusion. We have changed to ‘PrAoI’ throughout the manuscript.

4) As mentioned above, it is of interest to know whether patients were informed of the results of their throat culture and offered antibacterial treatment if they tested positive for beta hemolytic streptococcus.

All patients with group A beta-hemolytic streptococcal infection on throat culture were treated with antibiotics. Patients with group C streptococcal infection who had not improved by the time of the culture report (usually by 48
hours) were also treated with antibiotics.

5) Related to the section on Safety, pages 11-12, the Consort flow diagram indicates one patient left the active treatment arm of the trial due to an AE – what was that?

The AE was headache. This information has been added as a footnote to the CONSORT flow diagram.

6) In the Discussion, page 13, the authors describe a "major limitation" of the study as follows: "Approximately 40% of the patients reported the onset of throat symptoms in the previous 3 days (and five patients were inadvertently admitted with onset in the previous 4–5 days). These patients tended to dilute the differentiation between active medication and placebo as their throat symptoms improved naturally over the first 24 hours of the study (which was actually the 4th, 5th, or 6th day of these patients’ symptoms).” However, no results or statistical analyses are provided to support this observation about dilution -- please do so or remove this assertion.

We have removed this assumption, as suggested. However, we do feel it is important to highlight that the first 24 hours of the study was actually the 4th, 5th, or 6th day of many patients’ symptoms. We consider this to be a study limitation because we have evaluated symptom relief in patients at the mid–end stages of their sore throat episode rather than the early stages, when symptoms are usually worst.

ANOVA for the study’s primary endpoint:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>1</td>
<td>188</td>
<td>9.72</td>
<td>0.0021</td>
</tr>
<tr>
<td>Site</td>
<td>3</td>
<td>188</td>
<td>0.14</td>
<td>0.9376</td>
</tr>
<tr>
<td>Baseline STPIS</td>
<td>1</td>
<td>188</td>
<td>0</td>
<td>0.9877</td>
</tr>
</tbody>
</table>