Reviewer's report

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

Version: 2
Date: 1 December 2013
Reviewer: Christian de Mey

Reviewer's report:

The authors describe the pain relieving effects of single dose and repeated doses of a lozenge containing 8.75 mg flurbiprofen relative to sucking a similarly flavoured placebo lozenge investigated in randomised, placebo-controlled, double-blind parallel-group fashion in patients with acute sore throat of moderate-to-severe pharyngitis. In their approach, the investigators impose following methodological restrictions: a) enrolment was confined to patients with pharyngitis as the predominant URTI-condition, b) patients had to have had an onset of (possibly even very mild) symptoms within the last three days, c) baseline pain intensity had to be sufficiently high at baseline, and d) efficacy was to be evaluated not only in terms of pain relief, but also with regard to feeling “swollen throat” and “difficulty swallowing”.

Patients were observed on-site over the first 2 hours after a first dose of the medication and were then released from the clinic; they were provided with sufficient supplies and precise instructions to use the assigned medication every 3-6 hours with up to 5 doses per day. They also were provided with acetaminophen as rescue medication. Efficacy scores (STPIS, DSS, SwoTS) were recorded at baseline and then 1 and 2 hours after the first dose, every hour for the first 24 hours after the first dose (when awake), and then 1 and 2 hours after each dose taken over days D2-D7. The primary criterion was the time-weighted summed difference in pain intensity over the first 24 hours (SPID24). Criteria were contrasted between the treatment groups by ANOVA with treatment and site as fixed effects and pre-treatment baseline intensity as covariate. Possible confounding by the use of rescue medication was resolved by imputing the baseline scores for all subsequent measurement points (BOCF). A similar approach was taken for the percentage of patients who reported at least 20% reduction from baseline pain intensity over the first 6 hours after the first lozenge.

Recommendations

Revision of the manuscript is recommended with regard to the following:

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 …! »).

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

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.

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.

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 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?

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).

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 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?

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests