Scandinavian Journal of Pain 4 (2013) 46–47

Contents lists available at SciVerse ScienceDirect

Scandinavian Journal of Pain

journal homepage: www.ScandinavianJournalPain.com

Editorial comment

Validity of conclusions on treatment efficacy: Difficulties in patient recruitment and a large number of drop-outs may lead to bias

Eva Skovlund

Norwegian Institute for Public Health and School of Pharmacy, University of Oslo, Oslo, Norway

Hougaard et al. [1] in this issue of the *Scandinavian Journal of Pain* report on a randomised, double blind, placebo controlled cross-over study in the aura phase of migraine from which we can learn a great deal about trial design and generalisability of trial results.

NXN-188 is an oral dual action nNOS-inhibitor and $5HT_{1B/1D}$ receptor agonist. A total of 50 patients were randomised between two different sequences of NXN-188 and placebo. However, only 18 patients completed the trial and were available for the planned analysis. No statistically significant difference between NXN-188 and placebo was demonstrated. As stated by the authors, the low number of patients completing the study prevents a firm conclusion on lack of efficacy. Nevertheless, the results presented do not seem very promising.

The authors have experienced large problems recruiting patients to the trial and have faced an unexpectedly high proportion of drop-outs before the second treatment period. Probably they would have been hesitant to initiate this trial had they known that the final statistical analysis would include only 18 completers. The figure illustrating patient disposition illustrates their challenge. Out of 615 patients contacted by telephone 563 did not meet entry criteria or did not wish to participate in the study. Only 52 patients (8.5%) came for a screening visit out of which 50 were eligible and were randomised between two different sequences. It turned out that 17 patients (34%) did not treat an attack within 16 weeks, leaving 33 patients receiving their first dose (66% of those randomised). Eleven patients did not treat a new attack within 16 weeks after the first dose and were excluded according to the protocol, and 4 were lost to follow-up or did not meet criteria for drug treatment, leaving only 18 completers (36% of randomised patients).

1. Sample size and power to detect a treatment effect

According to the authors a trial including 33 participants would have 80% power to detect a difference in efficacy of 35% (60% on NXN-188 and 25% on placebo). With only 18 completers, the power is substantially reduced. A willing patient originally excluded due to protocol violation was re-included. However understandable when recruitment problems are faced such practice is not recommended.

The value of post hoc power calculation is of course very limited, but a few comments on sample size need to be made. Unfortunately the above information is insufficient for appropriate sample size estimation. With a paired design one also needs to make assumptions on the proportion of concordant and discordant pairs of observations from the same patient. In practice, with 18 patients contributing two observations each, the power could be as low as 40% if only few patients score identically on both treatments. The power is around 70% given that only one patient (out of the expected 25% responders on placebo) responds to placebo but not to the active treatment. With such a small patient number, a large *p*-value and thus an inconclusive result cannot be regarded as evidence of no effect of the active treatment.

The low power is not the only cause for concern. It is also questionable whether the patient sample included in the statistical analysis can be regarded as representative and whether the study results can be generalised.

2. Patient selection operates on different levels

The first level is the recruitment of patients to the trial. Patient selection leading to potential lack of external validity of findings is a general problem associated with all randomised clinical trials. Whether or not a common problem, it is certainly worthwhile trying to understand why more than 90% of migraine patients contacted were either not eligible or not willing to participate in the NXN-188 trial.

It is well known that generalisability may be reduced if a large proportion of patients seen in daily practice are ineligible or if many eligible patients are not included. Patients willing to be entered into a clinical trial often differ from those who refuse randomisation, and the pre-treatment characteristics of randomised and non-randomised patients may differ substantially with respect to clinical variables that may affect outcome [2,3].

Perhaps less strict inclusion- and exclusion-criteria could have reduced the number of ineligible patients and thereby increased trial feasibility.

A large proportion of eligible patients not willing to be randomised could however still affect generalisability of results.

DOI of refers to article: http://dx.doi.org/10.1016/j.sjpain.2012.08.002.
E-mail address: eva.skovlund@farmasi.uio.no

1877-8860/$ – see front matter
http://dx.doi.org/10.1016/j.sjpain.2012.09.006

3. Drop-outs and analysis of completers only

A more worrying level of selection potentially leading to bias is withdrawal of randomised patients during the study period (drop-outs). At this level not only external validity may be questioned, but even the internal validity of the trial is at stake. It may be reasonable to assume that the majority of those 15 patients who treated a first attack, but never the second one, actually did not experience an attack within the next 16 weeks.

But the case could also be that they chose not to comply due to their experience during the first treatment period. If that is the case, the validity of the analysis of completers is questionable. In any case the second observation is hardly missing at random and an analysis of completers only is potentially biased [4].

4. A parallel group design superior to cross-over design with many drop-outs

Even if a larger number of patients would have been needed a parallel group design would reduce drop-outs and might have turned out to be a more feasible option.

5. Need to plan for valid conclusions

Some clinical trials fail in the sense that the experimental treatment is not efficacious. However disappointing that may be, a bigger problem is clinical trials failing due to lack of power or inability to preclude biased efficacy estimates. Whenever a trial is planned we need to ensure that a valid conclusion can be reached. Administering an experimental treatment to patients in a clinical trial should be regarded as unethical if an inconclusive result is to be expected.

6. Assess feasibility

Whatever the reasons for the small proportion of completers in the NXN-188 trial the experience calls for realistic assessment of eligibility criteria and feasibility of study design in similar planned projects. A small pilot study in the planning phase followed by close monitoring and frequent contact with patients during the trial may prove invaluable not only in migraine trials. That could prevent spending a lot of research resources on trials not leading to reasonably precise efficacy estimates and firm conclusions.

References

[1] Hougaard A, Hauge AW, Guo S, Tfelt-Hansen P. The nitric oxide synthase inhibitor and serotonin-receptor agonist NXN-188 during the aura phase of migraine with aura: A randomized, double-blind, placebo-controlled cross-over study. Scand J Pain 2013;4:48–52.

[2] Moynihan C, Bliss JM, Davidson J, Burchell L, Horwich A. Evaluation of adjuvant psychological therapy in patients with testicular cancer: randomised controlled trial. Br Med J 1998;316:429–35.

[3] Licht RW, Gouliaev G, Vestergaard P, Frydenberg M. Generalisability of results from randomised drug trials. A trial on antimanic treatment. Br J Psychiatry 1997;170:264–7.

[4] ICH Topic E9. Statistical principles for clinical trials. <http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002928.pdf>.