Scandinavian Journal of Pain 4 (2013) 217–219

Contents lists available at ScienceDirect

Scandinavian Journal of Pain

journal homepage: www.ScandinavianJournalPain.com

Editorial comment

Statistical pearls: Importance of effect-size, blinding, randomization, publication bias, and the overestimated *p*-values

Harald Breivik[a,b,c,*], Leiv Arne Rosseland[a,c], Audun Stubhaug[b,c]

[a] Department of Anaesthesiology, Division of Emergencies and Critical Care, Oslo University Hospital – Rikshospitalet, Oslo, Norway
[b] Department of Pain Management and Research, Rikshospitalet, Oslo University Hospital, Oslo, Norway
[c] Faculty of Medicine, University of Oslo, Oslo, Norway

In this issue of the *Scandinavian Journal of Pain*, our associate editor for medical statistics, Professor Eva Skovlund focuses on some of the important, and frequently misunderstood, basic issues of statistical analyses of medical research data [1].

1. The over-rated *p*-value is often misunderstood

p-values are always there in scientific publications from medical researchers. They impress reviewers, editors, and readers alike, more than they deserve. A low *p*-value is taken as a "proof" of the truth of the research-question, but the *p*-value only indicates the likelihood of finding the observed difference between treatments by random/by chance, even if there is no difference. Still, the observed difference can be of limited clinical relevance, in spite of a low *p*-value [1].

2. Effect-size and clinical significance

As Eva Skovlund writes in her review [1]: "a *p*-value does neither assess the size of an effect as such nor whether a statistically significant result is of any clinical relevance".

For clinically meaningful information from the statistical analyses, the size of the effect and its 95% confidence interval are needed.

The effect-size is the difference between the observed value(s) and the expected value(s). If the null-hypothesis is true, i.e. that there is no difference between the therapies, the effect-size will be zero. Therefore, the effect-size should always be included in clinical trial reports [1].

It is important that the observed effect size is valid (only) for the population studied, with its baseline severity of illness, such as pain intensity. Thus, the results and the effect size may not be the same in a different study population or in a population with less or more severe illness (pain).

3. Pre-study estimate of sample size

The clinician knows the size of the effect that is interesting and relevant for the new intervention compared with standard treatment or placebo. With this knowledge, it is possible to estimate the number of patients needed in a trial in order to find such a difference in effect with reasonable certainty. A power of about 80 will ensure that if the researcher repeats the trial 10 times, a statistically significant difference will be detected in 8 of the 10 trials.

It is not always easy to define one primary outcome-measurement on which value the sample-size estimate should be based. Sometimes the outcome of interest is repeated measurements of a variable. Statistical methods based on regression or multi level modelling, have become more common in clinical pain trials [2,3]. Multi level modelling is powerful, but sample size calculations are more complicated because one has to take into consideration the correlation within the participating individuals.

4. Ethical aspects of planning clinical trials

It is ethically unacceptable to expose patients to possible risks of new interventions in a clinical trial of insufficient size. A trial unable to detect a real difference between treatment groups due to poor planning and group-size calculation is not only unethical but also a waste of resources. An Ethical Research Board cannot approve a clinical trial protocol that has no estimate of the sample size that is needed to find a difference between treatments if in fact a clinically important difference is expected. We agree with Eva Skovlund that an Ethical Research Board also should evaluate the planned statistical methods [1]. Missing values are the rule in clinical research, and she describes clearly important risks of biased analyses because of missing values [1]. This loss of validity of outcome-estimates can, in addition to the strategies outlined by Eva Skovlund, be reduced by using multi level modelling since all participants will be included, even if some values are missing [2].

Another ethical issue in clinical pain intervention trials is the patients' baseline pain intensity. It is a striking observation that numerous pain studies attempt to measure pain relief when there is no certainty about the level of pain that the treatment is expected to reduce. A trial giving pain-relieving medication before surgery

DOI of refers to article: http://dx.doi.org/10.1016/j.sjpain.2013.07.023.
* Corresponding author at: Department of Pain Management and Research, Rikshospitalet, Oslo University Hospital, PB 4950 Nydalen, 0424 Oslo, Norway. Tel.: +47 23073691; fax: +47 23073690; mobile: +47 95865323.
E-mail address: harald.breivik@medisin.uio.no (H. Breivik).

http://dx.doi.org/10.1016/j.sjpain.2013.08.001

for postoperative pain is bound to include many patients who in fact do not need any pain medication after surgery. It is documented that including only patients with relevant (high enough) baseline pain-intensity will reduce the variance, and because the necessary sample size is directly related to the variance [1] the necessary sample size will be much smaller [4]. A good example of the importance of including only patients with sufficient baseline pain is the series of trials of pain relief from intra-articular morphine administered after knee arthroscopy only when the patients reported postoperative pain [5]. All other such studies had administered the morphine (or placebo) into the knee-joint at wound-closure to all patients and therefore included a majority of patients with no pain or only minor pain after arthroscopic procedures (see next paragraph) [3].

5. Publication bias is a major problem, and false medical "truths" based on publication bias are difficult to erase

Small trials that by chance, find spurious positive ($p < 0.05$!) outcome-differences are more often accepted by reviewers and editors of scientific journals than "negative" outcome studies. This *publication bias* has a tendency to grow and cause false medical "truths". An illustrative example is the result of many small studies falsely concluding that a dose of morphine injected into the knee joint at the end of the procedure caused prolonged postoperative analgesia. Editors, who believed that this was a fact, accepted more such false positive trial reports for publication. Researchers, who did not find positive outcome, thought maybe that they had done something wrong in their trial, and they did not bother to submit a trial report. Alternatively, if they did submit a manuscript, their trial reports were likely to be rejected by reviewers and editors.

In a systematic review where scientific quality and group size were emphasized, we found no evidence of analgesic effect of saline with 0.5–5 mg morphine into the knee joint, compared with saline alone [3]. Interestingly, a new meta-analysis (August 2013) concludes that there might be a positive effect [6]. However, the number of well designed trials with sufficient sample size is almost the same as 10 years ago and only one trial included patients with sufficient baseline pain [7]. This illustrates how difficult it is to remove a "false" medical truth, created by small-sample trials with large type I errors and publication bias – see also Hopewell et al. [15].

6. Obligations of journal editors to publish "negative" study-reports

This is one reason that the editors of the Scandinavian Journal of Pain accept for publication well planned, well conducted, and correctly analyzed trials with "negative" findings, i.e. they have to accept the truth of the null-hypothesis [8,9]. They can be even more important than trials that reject their null-hypothesis and report a positive trial outcome.

7. Blinding is necessary for documenting *specific* treatment effect

The context-sensitive therapeutic effect (often called a "placebo"-effect), is a powerful positive effect when patients with a long-lasting pain-condition are treated well in a team of dedicated nurses, physiotherapists, and pain doctors ("multidisciplinary pain management"). This context-sensitive effect goes on and on, as long as patients and treatment providers are convinced that what is being done is helpful (see also Breivik et al. [14]).

Comparing baseline-data with outcome-data can thus give low *p*-values, even if the treatment effect is mostly due to unspecific (but good) context-sensitive effects. This is the conclusion of Stephen Morley when estimating outcomes of cognitive-behavioural-therapy (CBT) for chronic pain [10].

Therefore, to be able to show any specific pain-relieving effect of an intervention or a new drug, it is mandatory to have watertight blinding of patients as well as the team involved in outcome assessment. If possible, all treatment providers should also be blinded.

8. Randomization, when done correctly, reduces selection bias

Conscious or unconscious *selection bias* can be reduced by randomization of patients to standard treatment (or placebo) and new treatment. However, there are "true" randomizations and there are "pseudo"-randomizations. An example of pseudo-randomization, that still occurs, is to give standard treatment on even-numbered dates, new treatment on odd-numbered dates [11]. This means that the researchers selecting candidates for inclusion (consciously or unconsciously) may include "good-responders" on days when the new treatment is given.

Descriptive statistics are needed for demographics and basic variables such as age, gender, weight etc. However, when *p*-values are printed for differences between groups, one may suspect that the authors do not trust their randomization-procedure [1].

9. Why placebo-treatment is ethically acceptable in pain trials

If two active treatments are compared and the tests find no difference between the treatment, this is not a proof of equivalence. First, we do not know if the study was able to find any difference at all. Including a placebo group allows the demonstration of study sensitivity when the analysis shows that active treatment differs from placebo. Secondly, the difference between a gold standard and placebo can serve as a yardstick in the evaluation of the new treatment [12,13].

Patients, who do not respond, will soon enough receive a rescue analgesic. Therefore, it is good ethical trial practice to include placebo-treatment in pain studies.

10. More statistical pearls made easy

In her important review Eva Skovlund explains several more important statistical issues such as *non-inferiority trials, multiple testing, what to do with missing values, intention to treat-analyses, and more* [1]. This review should be an obligatory piece for all novice clinical researchers. We are sure the many-seasoned clinical researcher also will benefit from reading this paper.

Conflict of interest

No conflict of interests declared.

References

[1] Skovlund E. Significance tests in clinical research – challenges and pitfalls. Scand J Pain 2013;4:220–3.
[2] Ma Ys, Mazumdar M, Memtsoudis SG. Beyond repeated-measures analysis of variance: advanced statistical methods for the analysis of longitudinal data in anesthesia research. Reg Anesth Pain Med 2012;37:99–105.
[3] Rosseland LA. No evidence for analgesic effect of intra-articular morphine after knee arthroscopy: a qualitative systematic review. Reg Anesth Pain Med 2005;30:83–98.
[4] Breivik EK, Haanaes HR, Barkvoll P. Upside assay sensitivity in a dental pain model. Eur J Pain 1998;2:179–86.
[5] Rosseland LA, Stubhaug A, Sandberg L, Breivik H. Intra-articular (IA) catheter administration of postoperative analgesics. A new trial design allows evaluation of baseline pain, demonstrates large variation in need of analgesics, and finds no analgesic effect of IA ketamine compared with IA saline. Pain 2003;104:25–34.

[6] Chao Zeng C, Gao S, Cheng L, Luo W, Li Y, Tu M, Tian J, Xu M, Zhang F, Jiang W, Wei L, Lei G. Single-dose intra-articular morphine after arthroscopic knee surgery: a meta-analysis of randomized placebo-controlled studies. Arthroscopy 2013;1:1–11.
[7] Solheim N, Rosseland LA, Stubhaug A. Intra-articular morphine 5 mg after knee arthroscopy does not produce significant pain relief when administered to patients with moderate to severe pain via an intra-articular catheter. Reg Anesth Pain Med 2006;31:506–13.
[8] Kontinen V, Kalso E. Why we are proud to publish well-performed negative clinical studies? Scand J Pain 2013;4:15–6.
[9] Breivik H, Stubhaug A, Hals EKB, Rosseland LA. Why we publish negative studies prescriptions on how to do clinical pain trials well. Scand J Pain 2010;1: 98–9.
[10] Morley S. Efficacy and effectiveness of cognitive behaviour therapy for chronic pain: progress and some challenges. Pain 2011;152(Suppl. 1): S99–106.
[11] Bernard SA, Gray TW, Buist MD, Jones BM, Silvester W, Gutteridge G, Smith K. Treatment of comatose survivors of out-of-hospital cardiac. Arrest with induced hypothermia. N Engl J Med 2002;346:557–63.
[12] Stubhaug A, Breivik H. Clinical trials: acute and chronic pain. In: Breivik H, Campbell WI, Nicholas MK, editors. Clinical Pain Management – Practice and Procedures. 2nd ed. London: Hodder Arnold; 2008. p. 514–28.
[13] Breivik Hals EK. Clinical trials: dental pain. In: Breivik H, Campbell WI, Nicholas MK, editors. Clinical Pain Management – Practice and Procedures. 2nd ed. London: Hodder Arnold; 2008. p. 529–37.
[14] Breivik H, Ljossa TM, Stengaard-Pedersen K, Persson J, Aro H, Villumsen J, Tvinnemose D. A 6-months, randomised, placebo-controlled evaluation of efficacy and tolerability of a low-dose 7 day buprenorphine transdermal patch in osteoarthritis participants naive to potent opioids. Scand J Pain 2010;1:122–41.
[15] Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results (review). Cochrane Library 2009;(1).