SPINE Volume 29, Number 18, pp 2088–2089

Letters

In Response:

Re: Hägg O, Fritzell P. Letter. Spine 2003; 29: 1160–1.

Hägg and Fritzell correctly commented that our study had a small sample size.[1,2] We discussed this in our original publication.[1] Also, Hägg and Fritzell correctly commented that estimated standard deviation in our study was much larger than postulated in the preliminary calculation of study size. Consequently, we underestimated the sample size.

Power is important when a clinically relevant difference is not statistically significant

Statistical power calculations can be valuable in planning an experiment. Leading statisticians do not recommend postexperiment power calculations as a guide to interpreting tests with statistically nonsignificant results.[3] In other words, the postexperimental calculations made by Hägg and Fritzell are not recommended.[2] Their attempt of reducing the observed difference of minimal clinical difference to tossing a coin is simply not valid.

Higher observed power does not imply stronger evidence for a null hypothesis that is not rejected. It is pointless to perform power calculations for hypotheses outside of the confidence interval because the data have already told us that these are unlikely values.[3] We know that values inside the confidence interval are not refuted by the data. According to a recent publication in *Spine,* type II error occurs when a study fails to detect a predetermined magnitude of difference between treatments.[4] In our study, the observed difference of 2.3 may well be the accurate estimate and is much less than the predetermined difference of 10 that the study was designed to detect. As discussed in our article, the upper limit of the confidence interval means that we cannot refute the predetermined difference; it is just less likely to occur.

The first and most essential step in the design of any study is to clearly state the primary question or purpose. We clearly defined the primary outcome to determine the sample size and for proper statistical analysis. Hägg and Fritzell commented that an Oswestry Disability Index of 10 is probably a minimum to be clinically important.[2] In the study populations comprising the Swedish and the Norwegian study, the measurement error was estimated to 12.[5,9] This means that the observed difference of 2.3 in our study has no clinical relevance.

The sample mean is the best indicator of the population value

The lack of precision is reflected in the wide confidence interval (−6.7–11.4), which includes the difference that we considered clinically relevant. Regardless of the width of the confidence interval, the best indicator of the population value is the sample estimate of 2.3. Hägg and Fritzell calculated that 80% power in our study requires a doubled sample size.[2] It is possible to double the sample size in our study. The Swedish study included patients who had been previously operated for disc herniation, whereas we designed 2 separate studies. The second study was presented at the European League against Rheumatism in Lisbon in 2003, but has not yet been published.[6] The mean difference was −3.7 (−13.5–6.2). Thus, by mixing, the observed difference had been close to zero and the confidence interval had been smaller.

A 25% difference in treatment response is considered a medium relative difference.[4] The difference in improvement in our study was 17%, although the difference was not clinically significant. Percentages, therefore, may be misleading. When surgery is compared with nonsurgical treatment, differences need to be clinically significant.

The correct price is what the marked is willing to pay. We apologize for making incorrect referrals to the Swedish study. Costs were presented at a conference in Norway in 2001 but have not published until recently.[7] Fritzell et al. estimated hospital costs (1999 currency) within Swedish Public Health Care at 112,000 Swedish kroner or $14,500 US. The estimated costs are much less than the marked is willing to pay. In 2002, a populistic political decision approved 1 billion Norwegian kroner or $149 million US to give patients on a waiting list for surgery in Norway an operation abroad. A large number of lumbar fusions were performed at private spine centers in Sweden at an average prize of approximately 300,000 Norwegian kroner or $44,500 US. To our knowledge, prices for lumbar fusion in California exceed the price at Swedish private centers.

Fritzell et al. made a sensitivity analysis to show that surgery are more cost-effective if costs for rehabilitation is higher than observed, and if costs for the operation and family support is lower.[7] We agree that the marked price for nonoperative treatment may be higher if the marked believe that such treatment is as effective as surgery for chronic low back pain. In our opinion, Fritzell et al. underestimate the price for lumbar fusion in the current

health marked in Western societies. The cost for family support was based on retrospectively collected data. A questionnaire was sent to the patients after the 2-year follow up, probably several years after, but this is not reported in their article.[7] Based on this information, costs for noninstrumented and instrumented fusion is no longer significantly different. The validity of the retrospective telephone interviews and questionnaires are not discussed in their paper. The Swedish study is supported by the industry.

Was the difference in the Swedish study clinically significant?

The main difference between the Swedish and the Norwegian study is the improvement in the nonsurgical group. The improvements in ODI after surgery are comparable in the 2 studies, 15.6 in the Norwegian study and 11.6 (at 2 years) in the Swedish study. The improvement after nonsurgical treatment is 12.3 in the Norwegian study and 2.8 in the Swedish study. The difference between treatments in the Swedish study is 8.8, which is less than the measurement error. This means that despite a *P* value of 0.001, the clinical relevance of the difference is questionable.

Knowledge about the behavioral therapy approach in exercises was available in Sweden when the Swedish study was started. This approach was better compared with control subjects, which were treated by their general practitioners and given analgesics and unspecified physiotherapy in a randomized Swedish study published in 1992.[8] The Swedish Lumbar Spine Study Group should have acknowledged these results to offer the best nonsurgical treatment.

Despite the limitation of the Norwegian and the Swedish studies, we believe that both studies have contributed more to the evaluation of the clinical efficacy of lumbar fusion than previously published studies. We are waiting for the publication of a randomized, controlled trial that combines the advantages of the Norwegian and the Swedish study.

Jens Ivar Brox, MD, PhD
Olav Reikerås, MD, PhD
Roger Sørensen, MD
Rolf Riise, MD
Department of Orthopaedics
National Hospital
Oslo, Norway

Øystein Nygård, MD, PhD
Department of Neurosurgery
St. Olavs Hospital
Trondheim, Sweden

Tor Ingebrigtsen, MD, PhD
Department of Neurosurgery
UNN, Tromsø

Anne Keller, MD, PhD
Department of Physical Medicine and Rehabilitation
Ullevål University Hospital
Oslo, Norway

Aage Indahl, MD, PhD
Department of Physical Medicine and Rehabilitation
Coast Hospital
Stavern

Inger Holm, PT, PhD
Astrid Friis, PT
Anne Kathrine Koller, PT
Department of Physiotherapy
National Hospital
Oslo, Norway

Hege Eriksen, PSY, PhD
Department of Biological and Medical Psychology
Bergen, Norway

References

1. Brox JI, Sørensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* 2003;28:1913–21.
2. Hägg O, Fritzell P. Letter. *Spine* 2004;29:1160–1.
3. Hoenig JM, Heisley DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am Stat* 2001;55:19–23.
4. Bailey CS, Fisher CG, Dvorak MF. Type II error in the spine surgical literature. *Spine* 2004;29:1146–9.
5. Hägg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003;12:12–20.
6. Brox JI. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises for the postlaminectomy syndrome. *Ann Rheum Dis* 2003;62(suppl 1):229.
7. Fritzell P, Hägg O, Jonsson D, et al. Cost-effectiveness of lumbar fusion and nonsurgical treatment for chronic low back pain in the Swedish Lumbar Spine Study: A multicenter, randomised, controlled trial from the Swedish Lumbar Spine Group. *Spine* 2004;29:421–34.
8. Lindstrom I, Öhlund C, Eek C, et al. Mobility, strength, and fitness after a graded activity program for patients with subacute low back pain. *Spine* 1992;17:641–9.
9. Holm I, Friis A, Storheim K, et al. Measuring self-reported functional status and pain in patients with chronic low back pain by postal questionnaires. A reliability study. *Spine* 2003;28:828–33.