**LETTER TO THE EDITOR**

# Letter to the Editor

Nicolas Pinsault[1], Maïa Dolgopoloff[2], Albin Guillaud[1]

Dear Editors,

The recent article "Evaluation of the clinical effectiveness of microkinesitherapy in post-traumatic cervicalgia: A randomized, double-blinded clinical trial" is probably one of the most ambitious study, from a methodological point of view, published on this topic. Such studies are a necessary first step to move forward in manual therapies.

However, we have several concerns about the statistical analysis proposed and we would like to provide an alternative analysis of the raw data supplied by authors.

First, the proposed analysis focus on within-group differences (pre-post intervention). The authors show statistically significant improvement of the primary outcome pain assessed by visual analogue scale (VAS), and the secondary outcome flexion-extension amplitude in the microkinesitherapy (MK) group when results for patients in the control group show no significant difference. On this basis, the authors conclude on the efficacy of MK. However, as pointed out by Senn[1], such way of doing is inconsistent with the aim of randomized controlled clinical trials (RCTs). RCTs require direct comparisons *between* outcomes of *the different groups.* Gelman[2] and Nieuwenhuis[3] also pointed out that the difference between "significant" and "not significant" is not itself statistically significant. Consequently once again, when making a comparison between two effects, researchers should report the statistical significance of their difference rather than the difference between their significance levels.

When authors proposed between-group analyses, only post-intervention outcomes were compared and the baseline differences between groups were not took into account. However, in the present study, a boxplot suggests that VAS pre-intervention was smaller in the control group than in the MK group (see fig 1). Moreover, the choice of an independent t-test to compare post-intervention outcomes appears to us as an invalid approach considering the small sample sizes (leading to a non-normal distribution of the samples mean).

Last, the minimal clinical interesting difference (MCID) is nowhere mentioned, nor the expected variability of the corresponding effect. Hence, the required sample size to detect the effect with 80% power was not calculated upon clinical bases but upon statistical considerations: "*Sixty patients were planned to be included in the protocol. This number allows satisfying the conditions of validity for the reasonable use of the chosen statistical tool*." As a result, the statistical power is unknown and probably much less than 80%. The authors stated that "*the reduced number of participants could have resulted in limiting the statistical strength of the tests. However, the results show a level of significance much higher than the 5% threshold retained.*" We want to mention here that a low statistical power not only reduces chance of detecting a true effect, but also reduces the likelihood that a statistically significant result reflects a true effect, as Button[5] pointed out.

Considering (a) the study design (RCT), (b) the primary outcome properties (The VAS being only ordinal considering that a given change in one patient may be of different magnitude than the same apparent change in another Kersten[6]) avoiding parametric statistics for analysis.), (c) the between group difference in baseline VAS (fig. 1) and (d) the difference
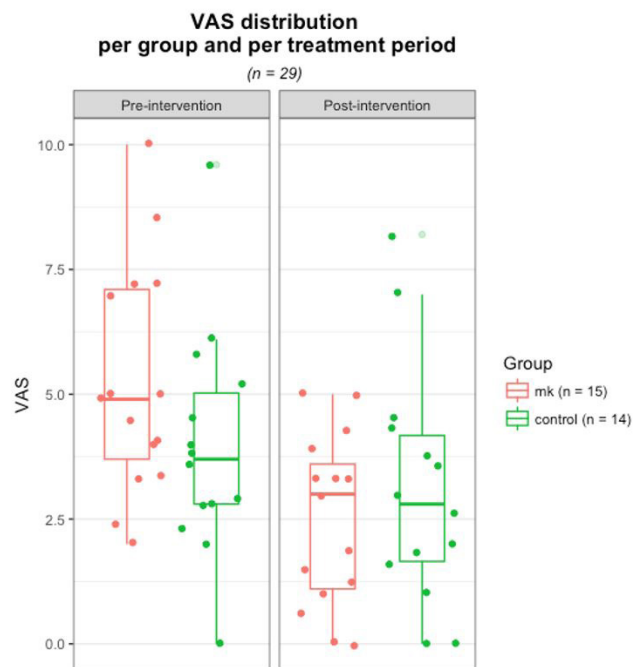


**Figure 1.** Visual analogue scale distribution per group and per treatment period. Note: VAS=Visual Analogue Scale; MK=Microkinesitherapy

**Correspondence:** Nicolas Pinsault. Université Grenoble Alpes. 621 avenue Centrale. 38400 Saint-Martin-d'Hères. +33 (0)4 57 42 21 42. nicolas.pinsault@univ-grenoble-alpes.fr

1. TIMC-IMAG UMR5525, Grenoble-Alpes University, Themas Team, Grenoble, France

Full list of author information is available at the end of the article.

in shapes of the VAS change score in each group (fig 2) invalidating the use of the independent Wilcoxon sum rank test for median comparison, none of the analyses proposed by authors is suitable for the estimation of the treatment effect. A proper statistical analysis would be to estimate difference in median VAS change scores (pre-post intervention) between groups using bootstraping approach. The corresponding estimated difference in medians (see additional information for R code) was -1.2 mm in favor of the MK group, and the associated bootstrap percentile interval at the 5% level was

[-4.0 ; 0.1] (fig 3). Thus, the difference in medians was not statistically significant.

For the secondary outcome (amplitude of flexion-extension), the estimated difference in medians was 10.5 degrees in favor of the MK group, and the associated bootstrap percentile interval at the 5% level was [-0.5 ; 20.5] (fig. 4). Thus, the difference in medians was also not statistically significant.

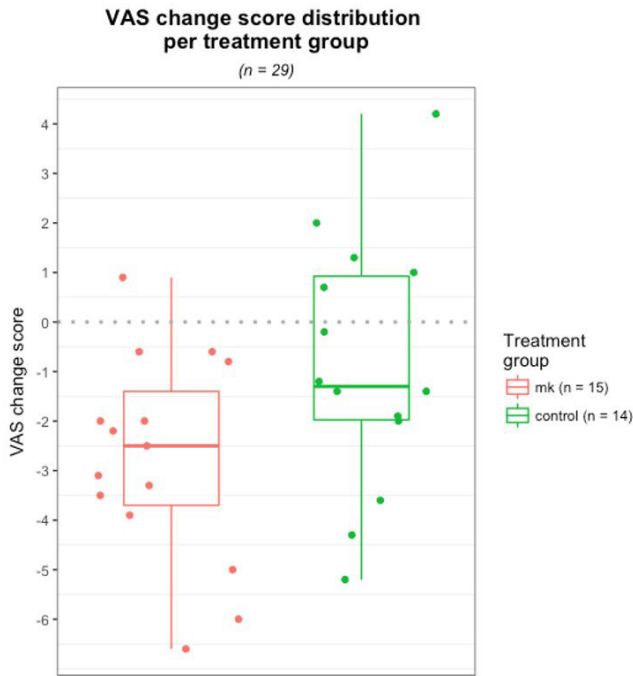In the context of unknown beta risk, it is impossible to conclude anything from these non-significant results.



**Figure 2.** Visual analogue scale change score distribution per treatment group. Note: VAS=Visual analogue scale; MK= microkinesitherapy.
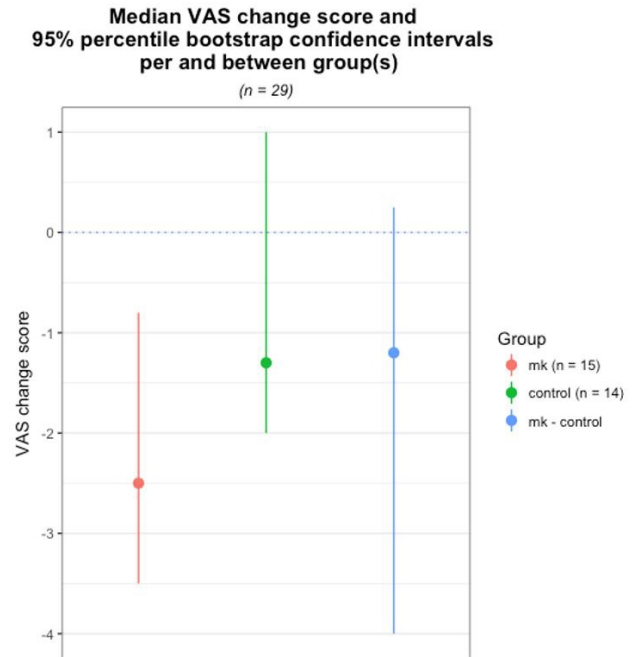


**Figure 3.** Median visual analogue scale and 95% percentile boostrap confidence intervals per and between group(s). Note: VAS=Visual analogue scale; MK= microkinesitherapy.
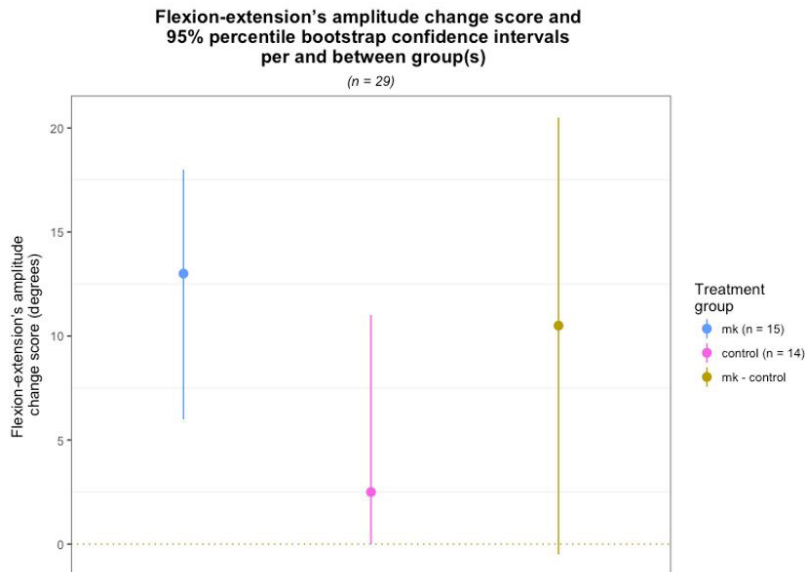


**Figure 4.** Flexio-extension's amplitude change score and 95% percentile boostrap confidence intervals per and between group(s). Note: MK=microkinesitherapy.

**AUTHOR'S CONTRIBUTION**
All the authos contributed for to prepare and write this letter.

**CONFLICTS OF INTEREST**
Nothing to declare

**AUTHOR DETAILS**
2Grenoble-Alpes University, Grenoble, France

## REFERENCES

1    SennS. Statistical Issues in Drug Development, 2nd Ed, 2007 John Wiley & Sons.

2    Gelman, A. & Stern, H. The difference between "significant" and "not significant" is not itself statistically significant. AmStat. 60, 328–331, 2006.

3    Nieuwenhuis, Forstmann, Wagenmakers. Erroneous analyses of interactions in neuroscience: a problem of significance. Nature Neuroscience 14(9):1105-7, 2011.

4    Hogg, Tanis (Probability and Statistical Inference, 7th Ed, Prentice Hall, 2006)

5    Button et al. «Power failure: why small sample size undermines the reliability of neuroscience » (Nature Reviews Neuroscience 14, 365-376, May 2013)

6    Kersten. The use of the visual analogue scale (vas) in rehabilitation outcomes.J Rehabil Med 2012; 44: 609–610