EDITORIAL

Editor-in-Chief

Nearly significant if only...

L A Harvey¹

© International Spinal Cord Society 2018

These are just some of the phrases that my red pen has traversed in recent months:

“nearly significant”

“trending towards significance”

“on the brink of significance”

“close to significant”

“approaching significance”

Such statements are often innocently inserted but the underlying suggestion is that the non-significant findings might be significant if the study was to be done again, or if the study had used a larger sample.

Perhaps statements like this largely originate from the mistaken belief that if the study was nearly significant this time then it will probably be significant next time. However, there is no basis to this belief. A p value does not get progressively smaller with replication. To the contrary, p values are random variables: if a study was to be exactly replicated many times the p value would jump around [1, 2]. This has been called the “dance of the p values” [2]. Even if the next study has a larger sample size than the first, there is no guarantee that a “nearly significant” result will become a “significant result”.

The “nearly significant” terminology may in part be due to the arbitrary nature of setting critical p values at 0.05. There is no good reason, beyond neatness, for setting the critical p value at 0.05 rather than 0.06 or 0.07. So

—surely—near enough is good enough. This logic might be acceptable if it were used consistently and in both directions: the researcher would need to be prepared to say just as many times that a p value of 0.04 was nearly insignificant or on the brink of insignificance as they were to say that a p value of 0.06 was nearly significant. Of course that is never going to become accepted practice! After all, a p value is a line drawn in the sand. Researchers are free to draw different lines before they start a study but they are not free to change the rules when their data do not oblige.

Spinal Cord will continue to remove any statements that imply that a p value is on the brink of, close to, or approaching significance because:

“...such descriptions give a misleading impression and undermine the principle of accurate reporting.”
(Pg. 1) [3]

We will also continue to strongly encourage authors to report point estimates and measures of uncertainty rather than p values [4, 5].

References

1. Motulsky H. Intuitive biostatistics: a nonmathematical guide to statistical thinking. Oxford, UK: Oxford University Press; 2014.
2. Cumming G. Dance of the p values. https://www.youtube.com/watch?feature=player_embedded&v=ez4DgdurRPg. Accessed on 16 October 2018.
3. Wood J, Freemantle N, King M, Nazareth I. Trap of trends to statistical significance: likelihood of near significant P value becoming more significant with extra data. *BMJ*. 2014;348:g2215.
4. Fidler F, Thomason N, Cumming G, Finch S, Leeman J. Editors can lead researchers to confidence intervals, but can't make them think: statistical reform lessons from medicine. *Psychol Sci*. 2004;15:119–26.
5. Harvey L. Statistical power calculations reflect our love affair with P -values and hypothesis testing: time for a fundamental change. *Spinal Cord*. 2014;52:2.

✉ L A Harvey
spinalcord@iscos.org.uk

¹ University of Sydney, Sydney, NSW, Australia