Reviewer's report

Title: An Evaluation of the Quality of Statistical Design and Analysis of Published Medical Research: Results from a Systematic Survey of General Orthopaedic Journals

Version: 1 Date: 27 February 2012

Reviewer: Erik Cobo

Reviewer's report:

This paper is well founded, easy to read and useful for readers.
I made some suggestions so that authors may consider them in improving the paper.

Major suggestions.

1) Please, to further disseminate your message, consider concentrating in major problems and to avoid relying in minor deficiencies or on aspects that may be considered an error by you but not to any statistician. (Remember: “to avoid statisticians governing your committee, just put them in an even number”). For example, from my point of view, the paper cited as correct (page 7, line 6) is still wrong, since making the average will decline its dispersion and the data will no longer pertain to the same population with the same variance. As another example, multiplicity is a very sensible topic (see BMJ 1998;316:1236–8). From my point of view, it is almost straightforward in the Neyman-Pearson decision framework to identify the hypothesis set that leads to the same decision (i.e, give permission to market a drug) and need adjustment. But from the Fisher evidence point of view (most medical papers), several different scientific questions may be addressed with the same data. So, assessing if authors fall into multiplicity can be a very difficult and sensible question. As a further example, I also prefer a parametric analysis after some transformation than a nonparametric one (pages 12 and 17), but only because it could provide a more readable effect size measure! I mean that this choice is, from my point of view, not a mistake, just free will. As a solution, please provide always references for your assessments. If possible, they should be on consensus documents such as reporting guidelines. To end this point, please consider concentrating your paper on main messages with solid documentation on agreement among statisticians and methodologists.

2) There is a lack of statistical inference tools to generalize the results of your ‘representative’ sample to the represented population. For any reader it would be useful to know the sampling uncertainty in making inferences. Please, seriously consider adding confidence intervals for your reported proportions.

Minor suggestions.

1) To facilitate reading, please consider adding tables or, better, graphics. For example, a Forest Plot for the main results.
2) Please, consider if ‘orphan p value’ is easy to be measured and worth commenting on. My guess is that most of your papers don’t report ‘effect size’ measures to address clinical relevance.

3) In page 6, paragraph ‘experimental unit’ needs a reference for the statement ‘…is a common error…’.

4) Please, consider avoiding vague expressions such as ‘some cases’, ‘for a number of’ or ‘others’, such as in lines 3-5 of page 8.

5) You already mentioned the different role of ‘sample size rationale’ in confirmatory and exploratory studies, which is reflected in the different wording in Consort (‘pre-determined’) and Strobe (‘arrived at’). So, please clarify what you mean by ‘justified the sample size’ in line 16, page 8.

6) Please consider clarifying if ‘data missing’ in line 6, page 9 refers to a complete unit, or just to a single value.

7) Please, note that the Wilcoxon and the Mann Whitney tests are so closely related that some authors employ the name ‘Wilcoxon’ to refer to both paired and unrelated samples (page 11, line 2).

8) Please, specify on page 11, line 6, that the case-control study was not paired.

9) Please, clarify why the sentence on page 12, line 3 is so wrong.

10) Please, consider removing ‘ethical’ in page 15, line 2. It should be understood that authors will delete any identification before publishing it. From my point of view, individual details belong to units, but results (with inference) belong to the population. What could be seen as unethical is keeping the anonymised data away from society.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

'I declare that I have no competing interests'