

Review of: "Mediumship for Pets: A Pilot Study With a Triple-Blind Protocol"

Romain-Daniel Gosselin¹

¹ Lausanne University Hospital, Lausanne, Switzerland

Potential competing interests: No potential competing interests to declare.

In their revised version of the manuscript, the authors add more quantitative (inferential) analysis. However, I really suggest they consult a statistician if they intend to keep revising their manuscript. So far, their methods make only little sense and are largely inappropriate, I'm afraid. I persist in my conclusion that no quantitative result in this work, with the presented design, suggests anything more than random variation in sampling. I would have many comments to make; I will restrict them to statistical priorities.

To start, not only do the authors persist in the unjustified use of one-tailed tests, but they also compare control pets' readings with human-intended readings, which makes no sense at all. I fail to understand the scientific logic here.

Next, I do not understand why they use different statistical approaches (Wilcoxon test vs. z-score) between Figures 1 and 3. The choice of a statistical procedure depends on the data structure (scale of the variables, number of groups...etc.) and some sets of assumptions. The data and population assumptions do not seem to differ between the two figures; the statistical methods should be the same (and the choice carefully justified).

Thirdly, the use of a z-score approach in Figure 3 is neither justified nor warranted; its use is likely to mislead the investigator with erroneous results. A z-score on proportions is possible when the sample size is large enough, leading the initial binomial data (due to the proportions) to tend to a normal distribution. Here, a group has a sample size of only $n=19$, which is largely insufficient. The strong unbalanced design between the 2 groups (pets' controls / human intended), although not formally posing a mathematical problem, may bias the inferential output. I strongly recommend performing a Chi-square test (or a Fisher Exact Test, or a G-test...) just like I did in my first review, which would evaluate the randomness of counts of observation rather than really comparing proportions.

With underpowered designs (like the "pets control" and "pets intended" groups), there is a tendency for the effect size to inflate for statistically significant results found (either false positives or true positives).

I will have to apologize, but I do not understand this result and discussion: "In the qualitative analysis, 18% of all correct information appeared to be obtained by a direct interaction with the pets. An identical analysis with deceased humans revealed a 12.6% of information of the same type."

There are many typos and odd sentences throughout the text that should be corrected.

