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

Romain-Daniel Gosselin[1]

1 Lausanne University Hospital, Lausanne, Switzerland

Potential competing interests: No potential competing interests to declare.

I feel compelled to admit the originality of the topic. I also praise the open data sharing.

However, I am intrigued by the mismatch between the experimental results and the apparent orientation of the authors' prose. The authors seem to suggest that their results could indicate that their triple blind protocol contributes to showing that pet mediumship works. Yet, the results essentially totally fail to show that pet mediumship convincingly works beyond experimental noise in the presented setting. The illustration is figure 3, which I re-analysed using a 2-tail Chi square test. It gives Chi square statistics of 0.3853 (p=0.5347) or 0.9418 (p=0.3318), depending on how we split the odd total counts of expected values in the contingency table.

When describing how the list of information (readings) was collected/coded, the authors describe, for instance: « For example, the sentence, "*I see a cat with brown fur and white spots, playing with a red ball launched by the sitter*", was formatted in a list of four different pieces of information: I see a cat; - Its fur is brown with white spots; - He plays with a red ball; - The ball is launched by a young boy ». I understand the objective here, but I also imagine the confounding factors. How specific are really these pieces of information? For example, what is the prevalence of brown and white cats in household cats? And in the sample of passed pets? How often would we randomly find cats who used to play with a red ball?... If we assume random chance that the medium gives this information (that is actually the null hypothesis) and if they are independent, did the authors dismiss or accept the entire statement about the pet if only one part (colour of the ball…) was wrong? If they really divided and split each piece of information from the complex sentences made by the medium, considering them totally independent, the probabilities of error/confounder is multiplied…

It is unfortunately a little easy to blame the lack of statistical power due to small sample size for the statistically non-significant results, especially when stating in the methods section that it was "expected." If it is foreseen during the design/data acquisition phase that the sample will be too small to reach a prespecified statistical power aiming to detect a prespecified effect/difference at a prespecified level of significance, then why conduct the study without collecting more data? A power/sample size estimation (whatever the method) should have been done prior to the data collection.

In terms of statistical analysis, the authors must clearly justify their use of one-tailed procedures (i.e., Student's t test) instead of two-tailed ones.

In the protocol, the authors describe that "The list of information (reading) related to the intended (requested) pet was sent to the sitter with a second list of information that served as a control, related to another similar pet, that is, a dog, cat, etc.,

labelled Reading A and Reading B.” They must explain how they did for the one squirrel that has no “similar pet” that could be used as a control. This must be clarified.

Which software did the authors use for statistical analysis?