

Review for “The WEIRD problem in a “non-WEIRD” context: A meta-research on the representativeness of human subjects in Chinese psychological research”

October 1st, 2022 (1st stage, round 2)

Reviewer: Kai Hiraishi (Keio University, Tokyo, Japan, kaihiraishi@keio.jp)

Thank you for revising the manuscript. I like the authors’ responses and the revisions. I have some minor comments and suggestions. Please see below.

Bayesian analysis

Thank you for explaining the planned Bayesian multinomial test. I think the explanation below is very informative. It helps me a lot to understand what the authors intend to test.

“The null hypothesis (H_0) is that the sample counts are generated by a specified set of population proportions. The alternative hypothesis (H_1) is that the sample counts are not generated by those population proportions.”

This means that in JASP, the alternative hypothesis for Bayesian multinomial test is an unconstrained alternative hypothesis. That is, H_0 is specified as a multinomial distribution with a specific set of parameters, but the alternative hypothesis H_a is simply not H_0 , i.e., all other possible parameters.

Then, I have few questions.

First, I am concerned about the relationship between the above explanation and the description on page 9 stated below.

The percentage data from one source is treated as the observed and the other is treated as expected. The null hypothesis (H_0) is that the observed percentage data are sampled from a multinomial distribution with parameters as defined by the expected percentage, the alternative hypothesis (H_a) is that the observed proportion data are sample from a multinomial distribution with equal probability for each cell.

Here, it seems that the definition of H_a is different from that in the first explanation (*i.e.*, H_a is not H_0). Could you give more explanation?

Second, I am concerned what will be the H_0 for the current study, especially for the first question.

The authors wrote that the first question is,

Firstly, whether the characteristics of Chinese samples reported in large-scale international collaborations are similar to those reported in Chinese psychological journals. (page 4)

Given the sentence, I have supposed that the sample characteristics of the studies reported in Chinese psychological journals will be the H_0 , and that of the international collaborative studies (hereafter big-team studies) will be the “sample count.” The Bayesian multinomial test will examine if the big-team study participants are sampled from the same (or similar) population as the participants reported in the Chinese psychological journals. However, the planned analysis is the other way around (on page 11 of the manuscript).

As I am unfamiliar with Bayesian analysis, I may be making a fundamental mistake. Nevertheless, it seems more natural to treat the big-team study data as the “sample count” if the authors are mainly concerned with the characteristics of the big-team samples. I suspect that many potential readers unfamiliar with Bayesian analysis would have a similar question. I would like the authors to explain in more detail why they set the big-team sample characteristics as H_0 .

Target population coding

I agree with the authors that representativeness has not been taken seriously in psychology. Given the situation, I like how the authors plan to code the target population, to code the explicitness of the target population description, and extract the exact sentences/words that describe the target population.

I agree with the authors that they can compare big-team and Chinese journal studies with the same target sample. I propose another set of analyses, comparing the big-team studies and the Chinese journal studies that lack an explicit description of their target population. In addition, the authors may compare those studies (big-team studies and Chinese journal studies combined) with the census and the family panel study.

It has been pointed out that, by not explicitly stating the target population, psychologists sometimes implicitly assume that their findings generalize to humans in general (Cheon et al., 2020; Kahalon et al., 2021). Thus I suppose that it is legitimate to assume that those studies, the studies that do not explicitly state the target population, should have collected representative samples that are representative of the general population. Of course, I do not expect it to be the case.

Updating the coding manual

I like the way the authors plan to revise and update their coding manual as below.

In the pre-coding stage, we first developed the initial version code manual based on the previous study (Arnett, 2008; Nielsen et al., 2017; Pollet & Saxton, 2019; Rad et al., 2018). Then, at least two coders will code ten random articles independently, they will compare the results, resolve the differences and revise the manual. After that, they will code another ten articles and compare the results and revise the coding manual again. This procedure will iterate until the disagreement between two coders is negligible. (page 7)

I have some minor questions regarding the coding manual¹ as below. I am confident that these will be addressed with the revision process.

1. The definition of “convenience sampling” in the sampling method section. Does it include crowdsourcing? Compared to standard random sampling, crowdsourcing is easier and more convenient. On the other hand, some crowdsourcing services such as prolific.co provide a “representative sample” that matches a nation’s population characteristics. While not random, they are expected to be more representative than the traditional convenience sampling via undergraduate psychology classes.
2. The classification of educational attainment. The classification consists of 1) lower than college and 2) college or higher. I wonder how coders should classify undergraduate students with the manual.
3. The classification of sample type. The authors provide only four classes; university students, students but not university students, infants and toddlers, and preschool children. Are they sufficient to cover the Chinese psychology participants?

I would suggest that further review on these points may not be necessary before stage 1 in-principle acceptance of the manuscript. However, if the authors wish a final check, or if the handing editor decides that a peer review based on the finalized manual is necessary, I am happy to help.

References

- Cheon, B. K., Melani, I., & Hong, Y.-Y. (2020). How USA-Centric Is Psychology? An Archival Study of Implicit Assumptions of Generalizability of Findings to Human Nature Based on Origins of Study Samples. *Social Psychological and Personality Science*, 11(7), 928–937. <https://doi.org/10.1177/1948550620927269>

¹ Chin_Subj_Code_Manual_V2.docx

Kahalon, R., Klein, V., Ksenofontov, I., Ullrich, J., & Wright, S. C. (2021). Mentioning the

Sample’s Country in the Article’s Title Leads to Bias in Research Evaluation. *Social*

Psychological and Personality Science, 19485506211024036.

<https://doi.org/10.1177/19485506211024036>