

Review of: "Immediate test-retest reliabilities of intention to quit smoking measures in current adult smokers"

Qihui Chen¹

¹ University of Minnesota

Potential competing interests: The author(s) declared that no potential competing interests exist.

Reviewer report on "Immediate test-retest reliabilities of intention to quit smoking measures in current adult smokers"

The paper under review examines an important issue in tobacco product perception and intention studies, that of the test-retest reliabilities of intention-to-quit-smoking measures, based on an online randomized, two-arm study. As survey respondents to survey questions might be affected by various factors such as their mood and home environment, test-retest reliability of key measures is key to the validity of the data collected. The paper is innovative in that it tackles an often-overlooked issue. While the design is simple, it's effective. The randomized study design also helps strengthen the internal validity of this study's findings. However, I have concerns over the external validity of this study, as discussed below. Below, I outline some areas needing improvements in the hope that my comments can help improve the paper.

1. Regarding external validity, more information on the selection of participants should be provided. One particular concern is that those who "indicated an interest in participating in market research studies" may not be comparable to those who did not indicate such an interest (and hence not in the database). This issue potentially undermines the external validity of the study (even if the study has strong internal validity). One way to strengthen the study's external validity is to compare basic demographic characteristics between the study sample and the general smoker population in the U.S., if possible. The authors also mentioned, "screened for inclusion using basic demographic information." It would be very helpful to provide more information on the screening criteria (—or was it simply "U.S. residents above 21 and smoking"?) and the process as part of the "quality control" procedure of the study. Also, any refusals? How high was the refusal rate? Any attritions between the test and re-test?

2. More details on "sample size calculation" are also needed to better inform the reader about why the two arms involved different numbers of participants. I understand one arm needs a large sample size to detect potential effects than the other arm, given their different contents, but how did you end up with those actual sample sizes needs more explanation.

3. It would be useful to show the specific content of the "filler task," at least in the Appendix. The test-retest reliabilities of the two measures may depend on the content and the length of the filler task. Understandably, it would not be difficult for healthy people to remember their answers provided a few minutes ago. So, the filler task serves to "distract" the

participants for a while, but whether such as distraction attempt is successful depends on the content and the length of the task.

4. Although the main findings are “almost perfect” and “perfect” between the two assessments, it would still be useful to break down the test results by demographic characteristics to see if some subgroup groups exhibit a unique pattern.

5. A related issue that this paper has no power to assess is: what happens if the time elapse between the test and re-test increases? For example, if you retest the participants after a few days, would their responses change dramatically. If possible, the authors might conduct a study in this direction.