Review of: "Population estimates of biomarkers of exposure to carbon monoxide, nicotine, and NNK in smokers and non-smokers"

Daniel Riggins[1]

1 Cook County Health and Hospitals System

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

Summary:
This study seeks to establish pooled population estimates for average levels of biomarkers of exposure related to cigarette smoking. They do so by compiling reported data in the literature and applying kernel density estimation to create estimated distributions. Their estimates show clinically and statistically significant differences between levels of each biomarker in current smokers vs non-smokers (and vs former-smokers). By doing so, they unify estimates from a variety of sources and hopefully boost validity and generalizability of these estimates over the findings of any individual study. These estimates will help compare the impact of interventions (like switching to potentially reduced risk tobacco products) on biomarker levels.

Assessment:
I appreciated the utility of this study in creating foundational data estimates for other work, and the overall approach of kernel density estimation is a valid one, however, further revisions should be applied to boost transparency around study selection.

Major Issue:
In the Literature Search section of the Results, the authors clearly state initial criteria for excluding 111 studies dropping from 195 total to 84 “meeting criteria”. However, the 84 studies meeting criteria further drop to 50 used for “pooling analysis inclusion”. The authors state the reasons for this second drop were documented in the supplementary tables, but I was unable to find this information while scanning through the tables. If they are indeed not there, please include this information and provide a summary of the reasons in the results section. If they are there, please demarcate more clearly for the ease of other reviewers.

Minor Issues:

- I would appreciate if the shape of the estimated distributions was more clearly described in the figures. See this article for different potential strategies: https://www.cedricscherer.com/2021/06/06/visualizing-distributions-with-raincloud-plots-and-how-to-create-them-with-ggplot2/
- In the spirit of open science, I would love if the analysis code and data were included in supplementary materials so

that other researchers can reproduce your findings and build on your work. See the Turing Way's Guide for Reproducible Research: https://the-turing-way.netlify.app/reproducible-research/reproducible-research.html

Small Edits and Comments:

I annotated the article in context using the Hypothesis platform, which you can view at:

https://hyp.is/d_eHSMDBEey9G0s2uI0dMg/www.qeios.com/read/ZJJ66O