

Peer Review for Maria Cuellar

Amanda Luby

1 General Comments

This paper is in really good shape, which made it kind of hard to pick out problem areas to fix. You do a great job at explaining your process thoroughly, both in text and with figures. It's clear that you've put time and effort into the writing and it comes off as polished and professional. I couldn't find any glaring problems, and found it really easy to read and understand.

I think one area that could be improved on in this section is reminding the reader of the 'big picture'. Sometimes in learning about a small step that you describe, I couldn't keep track of where it fit in with the problem that you're solving. Most of my comments below are easy fixes to somewhat 'awkward' phrasing of things, and trying to point out some places where you could remind the reader where you're taking them.

2 Specific Comments, suggestions, and edits

- Page 4, last paragraph: I wonder if maybe it would be helpful to describe the heuristic process before introducing the four steps of the method? I was a little confused by the example not following the four previous steps explicitly.

Also, I'd change 'To describe this process, we refer to Figure 2' to 'This process is illustrated in Figure 2' to make the sentence and paragraph flow a little better

- Figure 2: I love it. It's a great graphic and you do a good job explaining it in the caption, but you could maybe take out the last sentence in the caption since you specify at the beginning that you're just illustrating for one cohort.
- Page 5, second paragraph: I think this sentence has too many ideas and not enough stress positions. I think if you split it up, and also describe the problem before you fix it, it would be easier to understand. Maybe something like:
'As people age, they may not remember the exact age they were when they first used a given drug. To avoid this issue, we ask about marijuana-ever usage by age 25. This is justified because 99 percent of the US population who used marijuana started doing so before the age of 25 (citation)'
- Figure 3: when I printed the paper in grayscale, I couldn't tell the difference between the lines. Maybe make it black and white, or change the lines to be different dots or something?
- Page 6, second paragraph: This seemed really small and insignificant compared to the previous paragraph's discussion. Since the fix remains the same, maybe you could

introduce it at the same time as the rest of the problems at the beginning of section 3.1? That would make it seem less of an afterthought and the reader could think about it while reading about the NSDUH weights

- Section 3.1.1: I don't think you ever define what a 'synthetic cohort' is - it might be helpful to put a sentence with an explicit definition.
- Page 7, sentence 1: This is another long sentence that's kind of hard to keep track of. I also don't love starting sentences with a citation. What about:
 'If the cohort sizes are sufficiently large and if the true means within each cohort have a sufficient amount of time variation, Verbeek and Nijman (1992) have shown that the effects of ignoring the fact that only a synthetic panel is available is small. '
- Page 7, sentence 2: maybe add something like 'which makes all cohorts large enough to use a synthetic cohort representation'
- Page 7, sentence 3: I'd use 'Due to' instead of 'Thanks'
- Page 8, first full paragraph: It was a little unclear to me whether you are talking about the marijuana question in particular or all drugs?
- Section 3.2: I think it would be helpful to add a paragraph reminding us of the 4 steps and how the subsections fit in
- Section 3.2.1, third paragraph: I think this sentence is a little confusing to read due to the use of the conditional tense (would be, become, etc.) I think writing it as
 'We chose 1955 to 1975 as our years of interest because before 1955 the sample sizes are smaller than 100 when unweighted, and after 1975 the birth year is too close to our survey period.'
- Equation 2: I'm not crazy about the summation notation. Could you use $i \in I$ and then define I as the population domain? Just to make the equation look 'nice'. (Similar thing for equation 4)