

Assignment 15

Megha Joshi

I am a quantitative researcher at a research firm working mostly with program and policy evaluation.

One aspect related to data ethics that I have been thinking a lot about is how power analyses, that we have to conduct for research grants, preclude our ability to look at impacts etc. on groups that may have smaller sample sizes. We mostly write moderator analyses etc. as descriptive and tend to lump groups with smaller samples size into an “other” category. There are information on certain groups that we never know because of the way research is done. The grant writing process impacts the kinds of questions we are asking.

My company has also recently been exploring AI and it’s use for speeding up tasks like analysis of open-ended responses on survey for example. The company has set up guardrails preventing us from uploading proprietary information on for example OpenAI chats. I have also been wondering about getting consent from participants if we want to use their data to explore AI.

I also work with a lot of education data and have to be careful to not disclose information if the sample size of a group is very small that readers can trace the data to a student (they can identify the student and know their test score for example).

In terms of reproducible research, I am not really sure how it solves data ethics issues. I thought of it more in terms of being clear about documenting protocol, research design and code. And, that the results are robust a different research team conducting the research. And, perhaps helps with mitigating an analysts personal decisions and how much that impacts the takeaways. I am not sure how it solves issues about anonymity, data use, what kinds of questions are being examine.