Twitter Thread by Dr. Kristin Briney



Dr. Kristin Briney
@KristinBriney



I want to add an addendum to this thread from the other day to show why publishing an n=1 is so bad. It's because I can likely identify and put a name to this student.

(I'm not going to do that here but I am going to show you how easy it is.)

> I'm angry at everything right now, including the new the latest problematic article from C&RL.
>
> So I think it's time we talked about C&RL's history of publishing identifying information in articles.
>
> (Yes, @LibSkrat I'm finally telling this story)https://t.co/JPCks3tohr
>
> — Dr. Kristin Briney (@KristinBriney) January 6, 2021

To do that, let's talk about the IPEDS data set. IPEDS is a US database that contains a range of information about US universities, such as enrollment, test scores, graduation rates, etc.

One notable data table shows graduated students by major and ethnicity.

(FYI, here is the IPEDS data: https://t.co/K4OwsyLLsE It's an open database so you can explore at your leisure.)

Back to the "Completions" table which shows ethnicity by major. This happens to line up with the n=1 from the offending article which identified a student by their ethnicity and major.

Sorry, the n=1 was year in school and ethnicity but I've now used the IPEDS data to find out their major. Linking datasets on minority populations is very very powerful.

At this point, I can do a little digging either by department "happy graduation" announcements or even the graduation program to winnow the list down to all of the majors for that degree for the year. Now it's a matter of figuring out which minority student is the n=1.

This may not be foolproof to get me the exact name of the student, but using only publicly available data I've gotten *really* close to identifying them.

Again, I'm not actually doing this to identify that n=1 student but rather I want to show you how easily it can be done. Give me 30 minutes and I will have a list of potential names, and if one name is not "white" I might actually have THE name.

This is the danger of publishing n=1. I CAN IDENTIFY THAT STUDENT FROM PUBLICLY AVAILABLE DATA. It's not that hard.

All that tells me is that they were in the study, not actually what they did at the library.

BUT it also tells me that you don't care enough to protect the identities of your minority students.

And it tells me that the journal and peer reviewers don't know enough about de-identification and re-identification to prevent a paper with n=1 to be published.

So if you were curious in my thread from the other day about why n=1 is so bad, this is why. I have enough information to identify a student in the study by name.

I would really really like this to be the last time I have to write a thread like this or do an exercise to potentially identify a student who in NO WAY deserves to be identified just because they use the library.

So let's never published n=1 ever again. Basically, if your n's are under 10, you should see a red flag and either group small n's together or obscure the data in some way (e.g. "n<20").

I'll end with the fact that none of this is okay. In any field but especially in a field that is supposed to protect patron's privacy. We need to do better. All of us. Researchers, writers, reviewers, editors, etc. to make sure this doesn't happen again.

And finally, thanks to @hedgielib and @IandPangurBan who suggested I add this information on exactly how easy it is to identify someone by name in a published example of n=1.