Twitter Thread by David Broockman

David Broockman
@dbroockman

How do you write a PAP in practice, especially in limited time?

Regardless of how you feel about @The_JOP's new policy, a lot of political scientists will be writing PAPs for the first time.

I wanted to share my experiences/thoughts having written many PAPs.

Thread follows..■

1. What counts as a PAP? In my view, all that makes a PAP a PAP is that it pre-specifies procedures for data analysis, ideally as precisely as possible.

Note what this does *not* include: lit review, theory, etc.

Some people like to write PAPs that include all this, but...

..I do not think this is necessary. In my view, as long as your PAP pre-specifies procedures for data analysis, it's a PAP, regardless of what else you may want to add to it. Add more if you want, but you don't need to imo.

That means PAPs can be super short/fast to write.

2. How do you do this practically?

Idea 1: For people getting started, I highly recommend https://t.co/EbxOOIHr8Z. Screenshot attached. You'll answer 6 questions about your analysis, each answer requires ~a few sentences. Should take you around 15 minutes. And that's it!

The tool will create a) an anonymized PDF with your answers you can upload with a paper submission and b) a blinded link (eg https://t.co/hTHH0rb4wg) you can include that verifies you pre-registered. And you're done!

There's many more options. Here's two more things I've done:

Idea 2: This is my preferred method for survey experiments: Get simulated data, write your analysis code for it, and pre-register this code.

In Qualtrics, use Generate Test Responses https://t.co/ojakoqxlKr to create fake data, then write your analysis code for this fake data.

Advantages:

1) You need to write this code eventually anyway! The PAP is taking you 0 extra minutes.

2) In my experience, this gives you new ideas/flags problems in advance.

3) When you get the real data back, code is already written - results are now instant. ■

Big thing: Must comment your code and name your variables *REALLY* well so other humans (reviewers) can read it.

I'll usually comment certain lines like this: "# Main hypothesis test" "# Secondary test" "# Will also plan to present this" etc.

I then upload code as PDF & raw .do/.R files (with identifying info removed) to a new project in https://t.co/RmhFFTHUAA using the "open-ended preregistration" feature. Instructions here: https://t.co/AaOwAcMNqm.

Net time cost: ~5 min (since doing work you'd have to do later)

Idea 3: Short doc verbally describing analysis plan. Eg, https://t.co/BR2iLSdXpz. I usually do this for field experiments bc the analysis is straightforward but there's a lot of DVs and covariates and decisions about what DVs to put into what indices & how that I want to pre-reg.

Animating spirit: What am I going to want to say is pre-specified?

One goal is for writeup one day to be able to call as many things "pre-reg'd" as possible, allaying reader concerns that results are due to fishing. There's so many "researcher degrees of freedom"--minimize them!

Another goal: convince my future self. If I know that I made certain decisions ex ante, I won't worry later "hmm, is knowing how the results turn out affecting my thinking about what analysis decision to make?"

PAPs can also provide a time capsule of your expectations at the design stage. This has been helpful to me in making me confront when my expectations are wrong, forcing me to change my mind or abandon cherished but incorrect hunches.

Idea 4: I don't write PAPs like this, but many prefer to essentially write up the theory and design section of the paper in advance. See, e.g., https://t.co/UeMpk9uBDI.

I wish I did this more - seems like it has great advantages in helping people think through costly projects carefully in advance. HOWEVER, I don't think it is the only model.

Pick the model that is right and practical for you! I strongly believe there is no one right way.

Having read many PAPs, my view of the #1 thing that is missing: adjustments for multiple hypothesis testing. If you pre-register 20 hypotheses, you should be adjusting your p-values.

I have some projects in progress where I pre-register how to do this. More should.

In others, I clearly denote one hypothesis as the primary test and the rest as secondary, adjust those.

Would be good to see more of this, imo.

Lots of issues to discuss remain & I expect norms to evolve and to realize there were flaws I hadn't thought of in how I do PAPs now.

So, don't take any of this as me saying what I believe is the one true path. However, I hope this helps provide some practical advice to folks who are new to this. (end)

Addendum: I think a common misconception about PAPs is that their goal is to try to show off how many results you can predict in advance. "We predicted it, and it was true!"

I don't think that is the purpose. Purpose is to reduce researcher degrees of freedom. That's it.