2019

Nov

USING A PRE-ANALYSIS PLAN TO ENHANCE CREDIBILITY OF RESEARCH FINDINGS: THE CASE OF DEFORESTATION AND MALARIA

Jonah Busch | Research Associate

When Odysseus of Greek legend had to navigate his ship past the deadly Sirens, whose alluring songs tempted sailors to deviate off course and crash on rocks, he instructed his crew to lash him to the ship's mast, stuff their ears with wax, and not veer an inch from their charted course.

Image available online: view image

Similarly, when we started research on a controversial topic, and we wanted our findings to be as credible as possible, we employed a pre-analysis plan. That is, before looking at our data, we wrote down all our methods in advance, time-stamped our plan, and published it online. And then we adhered to it in the face of tempting mid-stream course revisions.

In a companion blog, we discuss the results of our research on whether deforestation increases malaria, which were published November 13 in *World Development*. In this blog we explore a different topic—seven reflections on using a pre-analysis plan for the first time.

Our pre-analysis plan lashed us to the mast. Anyone who's done empirical work of this nature knows how tempting it can be to make methodological changes after results start coming in. Maybe a particular result doesn't "look right." What if I take this variable out, or put a squared term in, or disaggregate in any number of possible ways? Because of our pre-analysis plan, we didn't do any of this.

Of course, the main point is not the strictures themselves. Rather, it is to reassure readers, peer reviewers, and even ourselves that we're not trying out methods every-which-way and then selectively reporting some results while relegating other results to the virtual file cabinet, potentially biasing the results.

Our pre-analysis plan made us think more upfront about how to best model the human-natural system. Because we committed to our methods in advance, we had to think a lot upfront about how to characterize a complex human-natural system as accurately as possible (Figure 1). We chose hypotheses to test based on published epidemiological literature, as opposed to hypothesizing

after results are known. And we included only those co-variates that should directly affect malaria, rather than just throwing in plausible-seeming co-variates that were easily available or that others had included before.

This was a lot of work upfront, but probably saved us some work we might otherwise have to do later anyway when confronted with perplexing results or peer reviewer feedback. The drawback is that full specification of methods in advance is often "close to impossible."

Image available online: view image

Figure 1. Coupled human-natural system of deforestation and malaria.

We used a split sample to head off coding mistakes. Another drawback of specifying all methods in advance is nervousness that we'd make mistakes in our initial plan and not be able to correct them later. We were able to prevent one type of mistake—coding errors—by using a split sample.

Before running our code on the full set of data from 90 country-years, we first tested our code on two randomly chosen country-years. Then we made any necessary coding changes. Importantly, we were only looking to make coding changes based on bugs we may have found in the code; not modeling changes based on the results of the small preliminary sample.

We suspect a pre-analysis plan helped our paper get published more easily, but it's hard to know. In principle, whether a scientific paper merits publication should be based on the importance of the research question and the soundness of the methods. But in practice, it's notoriously more difficult to publish a null result, i.e., a finding that something *didn't* have an effect. And we had a null result.

We believe that the pre-analysis plan made it easier than it would otherwise be to publish a paper with null results. But it's tough to know for sure. We don't know what would have happened if we'd submitted positive findings, or null results without a pre-analysis plan.

The pre-analysis plan helped us distinguish new analyses as exploratory. One of the reputed benefits of a pre-analysis plan is to head off requests from referees for extensive robustness checks. In response to comments from referees, we added four new tests, which is on the light side for work of this nature. Importantly, we were transparent that these new analyses weren't part of our pre-analysis plan, but were added later. And that the results of these new analyses should be considered exploratory. That is, they didn't test ex ante hypotheses, but generated hypotheses that can be tested in future work.

The treatment of pre-analysis plans is still evolving. When we began our work, one prominent repository for pre-analysis plans—Registry for International Development Impact Evaluations (RIDIE)—only accepted plans for experimental work, rather than non-experimental studies such as ours. Later they did accept our pre-analysis plans, registered here and here. Non-experimental studies make up just 4% of pre-analysis plans registered with the American Economic Association (AEA) or Evidence in Governance and Politics (EGAP) repositories, according to a recent study.

There is no clear reason why non-experimental studies shouldn't also pre-register their hypotheses and methods. If anything, the need is even greater. It's great to see the field moving in this direction, as part of a larger trend toward transparent and reproducible social science research.

We would do it again. Not every research paper needs a pre-analysis plan. They are especially well-suited for research questions with seemingly simple research designs, e.g. finding a yes-or-no answer, or testing the magnitude of an effect size. In this case, the cons of a pre-analysis plan—more work upfront—were outweighed by the pros—credibility, suspected greater ability to publish a null-finding, and pushing forward the field in terms of transparency.

The question of whether deforestation increases malaria prevalence in humans is not just of academic interest; it has practical importance. If deforestation increases malaria rates, then forest protection could be justified as an important part of public health campaigns under some conditions. But if it doesn't, then malaria eradication resources would be better spent on proven interventions such as insecticide-treated bed nets and indoor spraying. And efforts to conserve tropical forests would do better to emphasize the many other benefits of forests, such as climate stabilization, biodiversity habitat, and regional rainfall production.

We hope that our use of a pre-analysis plan bolsters the credibility of our findings that deforestation hasn't increased malaria prevalence in Africa. We plan to use pre-analysis plans again, when appropriate, and we encourage other researchers to do the same.

Again, for more about the results of the paper itself, see this companion blog.