

GOODMAN & REED: A Friendly Debate about Pre-Registration

Background: Nat Goodman is generally pessimistic about the benefits of pre-registration. Bob Reed is generally optimistic about pre-registration. What follows is a back-and-forth dialogue about what each likes and dislikes about pre-registration.

[GOODMAN, Opening Statement] We need to remember that science is a contradictory gimish of activities: creative and mundane; fuelled by curiosity and dogged by methodological minutia; fascinating and boring; rigorous and intuitive; exploratory, iterative, incremental, and definitive.

Instead of trying to rein in the galloping horse we call science, we should be trying to spur it on. We should be looking for new methods that will make scientists more productive, able to produce results more quickly -- and yes, this means producing more bad results as well as more good.

[REED, Opening Statement] In a recent interview, Nicole Lazak, co-author of the [editorial](#) accompanying *The American Statistician's* special issue on statistical significance, identified pre-registration as one of the “best ways” forward for science (see [here](#)).

The hope is that pre-registration will provide some discipline on researchers' tendencies to “graze” through various research outputs in search of something interesting. It is precisely that kind of “grazing” that encourages the discovery of spurious relationships. As spurious relationships are the root cause of the “replication crisis”, pre-registration provides direct medicine for the sickness.

[GOODMAN] Pre-registration is important for confirmatory research but irrelevant for exploratory work. The purpose of pre-registration is to eliminate *post hoc* investigator bias. To accomplish this, preregistered protocols must fully specify the study (including data analysis) with sufficient detail that a completely different team could carry out the work. This may sound over-the-top but is the norm in clinical trials of new drugs and focused social science replication projects.

Many people support a soft form of pre-registration in which the preregistered protocol is simply a statement of intent and suggest that this form of pre-registration can be used for exploratory research. I don't see the point. In my experience, exploratory research never goes as expected; we learn how to do the study by doing the study. Comparing the final method with the original plan is humbling to say the least.

[REED] Effective pre-registration should be more than a statement of intent. It should clearly identify the goals of the research, the set of observations to be used, variables to be included in the analysis, and principles for modifying the analysis (e.g., criteria for eliminating outliers). The goal is to prevent HARKing and (possibly unconscious) p-hacking.

Let me explain why I believe pre-registration can be effective in preventing HARKing and reducing p-hacking.

A lot of research consists of looking for patterns in data. In other words, exploratory research. However, too often the patterns one observes are the results of random chance. This itself wouldn't be so bad if there was a feasible way to adjust the statistical analysis to account for all the paths one had taken through the garden. Instead, researchers report the results of their exploratory analysis as if it were the one-shot, statistical experiment that significance testing presumes.

Pre-registration limits the number of paths one explores, making it less likely that one stumbles upon a random-induced pattern. Where one discovers something after departing from the pre-registration plan, it helps readers to properly categorize the finding as exploratory, rather than confirmatory.

It is important to note that pre-registration does not preclude researchers from exploring data to look for interesting relationships. Rather, the goal of pre-registration is to get researchers to distinguish between confirmatory and exploratory findings when reporting their empirical results. In the former case, statistical inference is valid, assuming the researcher makes Bonferroni-type adjustments when conducting multiple tests. In the latter case, statistical inference is meaningless.

There is some evidence that it works! Recent studies report that effect sizes are smaller when studies have been pre-registered, and that there are fewer significant findings (see [here](#) and [here](#)).

[GOODMAN] There is also evidence that pre-registration has not worked. Early results from studies that have been pre-registered indicate that researchers have not been careful to distinguish exploratory from confirmatory results (see [here](#)). There is good reason to believe that these early returns are not aberrations.

According to your model, exploratory results should be viewed with greater scepticism than results from confirmatory analysis. But researchers who want to see their work published and have impact will always have an incentive to blur that distinction.

I am not alone in my pessimism about pre-registration. Others have also expressed concern that pre-registration does not address the problem of publication bias (see [here](#)).

Pre-registration is a non-cure for a misdiagnosed disease. Current scientific culture prizes hypothesis-driven research over exploratory work. But good hypotheses don't emerge full-blown from the imagination but rather derive from results of previous work, hopefully extended through imaginative speculation.

The reality is that the literature is filled with papers claiming to be "hypothesis-driven" but which are actually a composite of exploration, *post hoc* hypothesis generation, and weak confirmation. This is how science works. We should stop pretending otherwise.

Let me get back to what I think is a fundamental contradiction in pre-registration. As I understand it, economics research often involves analysis of pre-existing data. Since the data exists before the study begins, the only way to avoid post-hoc pattern discovery is to prevent the investigator from peeking at the data before pre-registering his research plan. This seems infeasible: how can someone design a study using a dataset without pretty deep knowledge of what's there?

[REED] It's not the peeking at the data which is the problem, it is estimating relationships. Suppose my data has one dependent variable, Y, and 10 possible explanatory variables, X1 to X10. Pre-registration is designed to reduce unrestricted foraging across all data combinations of X variables to find significant relationships with Y. It does this by requiring me to say in advance which relationships I will estimate. Yes, I must look at the data to see which variables are available and how many usable observations I have. No, this does not eliminate the value of a pre-registration plan.

[GOODMAN] Pre-registration puts the emphasis on the wrong thing. Instead, greater emphasis should be placed on developing confirmation procedures. Devising good confirmation procedures is an important area of methodological research. For example, in

machine learning the standard practice is to construct a model using one dataset and test the model on another dataset (if you have enough data) or through bootstrapping. This might just do the trick in fields like economics that depend on analysis of large preexisting databases.

Further, as others have noted, the “fast spread of pre-registration might in the end block” other approaches to solving problems of scientific reliability because “it might make people believe we have done enough” (see [here](#)).

[REED] I’m only somewhat familiar with the uses of bootstrapping, but I don’t think this can solve all problems related to p-hacking and HARKing. For example, if there is an omitted variable that is highly correlated with both an included variable and the dependent variable, the included variable will remain significant even if one bootstraps the sample. Thus, while these can be useful tools in the researcher’s toolbox, I don’t believe they are sufficiently powerful to preclude the use of other tools, like pre-registration.

With regard to pre-registration potentially crowding out more effective solutions, I agree this is a possibility, but I’d like to think that researchers could do the scientific equivalent of chewing gum and walking at the same time by adopting pre-registration + other things.

[REED, Conclusion] I think our “debate” has played out to the point of diminishing returns, so let me give my final spin on things. I think we both agree that pre-registration is not a silver bullet. First, we don’t want to tie researchers’ hands so they are prevented from exploring data. Second, pre-registration can be ignored and, worse, manipulated. These weaken its ameliorative potential. On these two points we both agree.

Where we disagree is that Nat thinks there is only a negligible benefit to pre-registration for exploratory research, while I think the benefit can be substantial. In my opinion, the benefit accrues mostly to well-intentioned researchers who might accidentally wander around the garden of forking paths without appreciating how it diminishes the significance of their findings (both statistical and practical). While this won’t eliminate the problem of p-hacking and HARKing, I think requiring researchers to complete a pre-analysis plan will make well-intentioned researchers less likely to fall into this trap. And if you believe that most researchers are well-intentioned, as I do, that can lead to a significant improvement in scientific practice, and reliability.