An Open Discussion on Promoting Transparency in Social Science Research

Mar. 20th, 2013   Send to Kindle

By Edward Miguel (Economics, UC Berkeley)

This CEGA Blog Forum builds on a seminal research meeting held at the University of California, Berkeley on December 7, 2012. The goal was to bring together a select interdisciplinary group of scholars – from biostatistics, economics, political science and psychology – with a shared interest in promoting transparency in empirical social science research.

There has been a flurry of activity regarding research transparency in recent years, within the academy and among research funders, driven by a recognition that too many influential research findings are fragile at best, if not entirely spurious or even fraudulent. But the increasingly heated debates on these critical issues have until now been "siloed" within individual academic disciplines, limiting their synergy and broader impacts. The December meeting (see presentations and discussions) drove home the point that there is a remarkable degree of commonality in the interests, goals and challenges facing scholars across the social science disciplines.

This inaugural CEGA Blog Forum aims to bring the fascinating conversations that took place at the Berkeley meeting to a wider audience, and to spark a public dialogue on these critical issues with the goal of clarifying the most productive ways forward. This is an especially timely debate, given: the American Economic Association's formal decision in 2012 to establish an online registry for

experimental studies; the new "design registry" established by the Experiments in Governance and Politics, or EGAP, group; serious discussion about a similar registry in the American Political Science Association's Experimental Research section; and the emergence of the Open Science Framework, developed by psychologists, as a plausible platform for registering pre-analysis plans and documenting other aspects of the research process. Yet there remains limited consensus regarding how exactly study registration will work in practice, and about the norms that could or should emerge around it. For example, is it possible – or even desirable – for all empirical social science studies to be registered? When and how should study registration be considered by funders and journals?

With my co-authors Kate Casey (of Stanford) and Rachel Glennerster (of MIT), I recently worked on transparency issues in empirical research, with a particular focus on how the use of pre-analysis plans (PAP's) can bolster the credibility of findings generated by experimental and other types of prospective studies. Our paper, published in the *Quarterly Journal of Economics* (see here), reports results from a study in Sierra Leone that estimated the impact of a community driven development (CDD) intervention on a wide range of local collective action and governance outcomes. What sets our study apart is our inclusion of a pre-analysis plan, which we registered with the Jameel Poverty Action Lab (J-PAL) hypothesis registry before analyzing the project's endline data. (Much of the discussion of the paper below builds directly on this joint work with Casey and Glennerster.)

How did study registration work for us in practice? In Casey, Glennerster and Miguel (2012) we discuss our experience with a pre-analysis plan in some detail, with the hope that sharing what we learned will contribute to the emerging debate on the pros and cons of PAP's in social science. The research and project teams agreed to a set of hypotheses regarding the likely areas of program impact in 2005 before the CDD intervention began. As the project came to a close in 2009, we fleshed out this document with the exact outcome measures and

econometric specifications we would use, based on the surveys that we had designed, and archived this pre-analysis plan.

Since as far back as Ed Leamer's important work in the 1970s, economists have recognized that "tying one's hands" with a pre-analysis plan is potentially useful in settings where researchers have wide discretion over what they report. Researchers may face professional incentives to affirm the priors of the academic discipline or the agenda of donors and policymakers; the latter is a critical issue these days for development economists who often work closely with implementing partners to study large-scale programs. Explicit *ex ante* agreements between researchers and program sponsors, like the one we had with our partners in Sierra Leone, can offer a layer of protection for "inconvenient" findings and thus reduce the scope for tendentious reporting. Adherence to a PAP thus reduces the risk of data mining or other selective presentation of empirical results ("cherry-picking") and generates correctly sized statistical tests, bolstering the credibility of the findings.

The interest in PAP's has grown with the recent spread of randomized evaluation methods in economics. While the experimental framework naturally imposes some narrowing of econometric specifications, there is still considerable flexibility for researchers to define the outcome measures of interest, group outcome variables into different hypothesis "families" or domains, identify population subgroups to test for heterogeneous effects, and include or exclude covariates. When there are a large number of plausible outcome measures of interest and when researchers plan to undertake subgroup analysis, PAP's are arguably particularly valuable. The process of writing a PAP may have the side benefit of forcing the researchers to more carefully think through their hypotheses beforehand, potentially improving the quality of the research design and data collection approach.

At the December 7, 2012 meeting in Berkeley, Kate Casey and Ben Olken of MIT offered stimulating presentations on these and a wider set of issues involved in the use of pre-analysis plans in economics based on their experience using them, and Aprajit Mahajan of UCLA provided thoughtful reactions to their arguments. I urge anyone interested in pre-analysis plans, study registration, and the broader issue of research transparency in social science to study these presentations. They were followed by a discussion of the existing clinical trials system, and new data adaptive analytical tools in statistics by Maya Petersen and Mark van der Laan, both of U.C. Berkeley, with reactions by Bryan Graham, also of Berkeley. The afternoon session featured detailed discussion about ongoing efforts to establish a trial registry within political science by Jeremy Weinstein and David Laitin, both of Stanford, with provocative discussions by Kevin Esterling of U.C. Riverside and Don Green of Columbia. The movement to carry out large numbers of study replications within psychology was surveyed by Brian Nosek of Virginia, who also gave a fascinating description of the Open Science Framework that he has pioneered. He was followed by Leif Nelson of U.C. Berkeley, who described the "p-hacking" approach to discovering selective reporting of results, with an entertaining discussion by Gabe Lenz also of U.C. Berkeley.

There was broad agreement among meeting participants that a system of registration for experimental trials would help round out the body of available research evidence, mitigating the publication bias that arises from underreporting null or counter-intuitive results. Together with a broader push toward research transparency along multiple dimensions – by making the sharing of data, analysis code, and research proposals standard practice, for instance – the registration of pre-analysis plans has the potential to improve the credibility of much empirical social science research.

Whether or not this promise is realized is another question, and one that we believe will require a concerted effort among scholars across the social sciences, in partnership with research funding organizations and journals, over multiple

years. We at the University of California's Center for Effective Global Action (CEGA) are launching a new effort, the Berkeley Initiative in Transparency in the Social Sciences, or BITSS, to help promote productive discussions on the critical and still unresolved issue of how best to promote research transparency in the social science. By facilitating public discussions like this CEGA Blog Forum, and by hosting regular meetings among a network of interested scholars (building on the successful December 2012 meeting), BITSS aims to build greater consensus around approaches that will enhance research transparency, and begin to address the pervasive problems that are so eloquently articulated in the other Forum contributions.

About the author: Edward Miguel is the Oxfam Professor in Environmental and Resource Economics and Faculty Director of the Center for Effective Global Action at the University of California, Berkeley, where he has taught since 2000. His main research focus is African economic development, including work on the economic causes and consequences of violence; the impact of ethnic divisions on local collective action; and interactions between health, education, environment, and productivity for the poor. He has conducted field work in Kenya, Sierra Leone, Tanzania, and India. Ted is a recipient of the Alfred P. Sloan Fellowship, the Kenneth J. Arrow Prize, and the Berkeley Distinguished Teaching Award. Miguel is author with Ray Fisman of *Economic Gangsters: Corruption,*

Violence and the Poverty of Nations (Princeton University Press 2008), and author of *Africa's Turn?* (MIT Press 2009)

This post is one of a ten-part series in which we ask researchers and experts to discuss transparency in empirical social science research across disciplines. It was initially published on CEGA blog on March 20, 2013. You can find the complete list of posts here.

---

https://www.bitss.org/an-open-discussion-on-promoting-transparency-in-social-science-research/