An Application of Pre-Analysis Plans (PAP) in Economics

Katherine Casey, Stanford GSB

CEGA Meeting on PAP and Study Registration in Social Science Research, November 2012

Internal Validity and Discretion

- There are well known fundamental concerns about data mining, specification search (Leamer 1974, 1983) and "cherry picking"
- While randomization naturally imposes some narrowing, substantial discretion remains to
  - Define the universe and measurement of outcomes
  - Group outcomes under hypotheses
  - Identify sub-group populations to test for heterogeneous effects
  - Choose econometric specifications and the covariate set
- Use and registry of pre-analysis plans remains relatively rare in economics, but gaining momentum

An Application in Economics

- Casey, Glennerster and Miguel (2012) estimates the impact of a community driven development (CDD) program in Sierra Leone on local public goods, institutional performance and social capital
- CDD aims to improve the capacity and performance of local governance and boost the inclusion of marginalized groups, like women and youth
- Large-scale randomized experiment with N=236 villages, a four year time-frame (2005-2009), detailed data collection
- We found strong positive effects on public goods and economic outcomes and none on institutions

Study Features That Posed Risks

- A primary object (social capital) only loosely defined
  - Imprecision gives donors unhappy with inconvenient results an "out" by claiming that the research simply measured the wrong outcomes
- Multifaceted and context specific subject (institutions)
  - Large number of relevant outcomes creates scope for fruitful cherry picking
  - Lack of standardized measures makes such tendentious reporting difficult to detect from the outside
- Several sub-groups of theoretical interest (including women, youth, war-affected, and poorest)
  - X sub-groups by Y hypotheses again invites cherry picking

The Paper We Could Have Written

Outcome Variable	Mean for controls	Treatment effect	Standard error	N	Hypo
	(1)	(2)	(3)	(4)	(5)
Panel A: GoBifo "Weakened" Institutions					
Attended meeting to decide what to do with the tarp	0.81	-0.04+	(0.02)	236	H5
Everybody had equal say in deciding how to use the tarp	0.51	-0.11+	(0.06)	232	H5
Community used the tarp (verified by physical assessment)	0.90	-0.08+	(0.04)	233	H4
Community can show research team the tarp	0.84	-0.12*	(0.05)	232	H5
Respondent would like to be a member of the VDC	0.36	-0.04*	(0.02)	236	H10
Respondent voted in the local government election (2008)	0.85	-0.04*	(0.02)	236	H10

The Paper We Could Have Written v2

Outcome Variable	Mean for controls	Treatment effect	Standard error	N	Hypo
	(1)	(2)	(3)	(4)	(5)

Panel B: GoBifo "Strengthened" Institutions

Outcome Variable	Mean for controls	Treatment effect	Standard error	N	Hypo
Community teachers have been trained	0.47	0.12+	(0.07)	173	H4
Respondent is a member of a women's group	0.24	0.06**	(0.02)	236	H8
Someone took minutes at the most recent community meeting	0.30	0.14*	(0.06)	227	H5
Building materials stored in a public place when not in use	0.13	0.25*	(0.10)	84	H5
Chiefdom official did not have the most influence over tarpaulin	0.54	0.06*	(0.03)	236	H6
Respondent agrees with "Responsible young people can be good leaders" and not "Only older people are mature enough to be	0.76	0.04*	(0.02)	236	H6, H12
Correctly able to name the year of the next general elections	0.19	0.04*	(0.02)	236	H9

Plan Timing and Content

- What is the optimal timing?
  - Pre-program versus pre-endline analysis
  - A hybrid approach: hypothesis document agreed with project teams in Oct 2005 before baseline data collected; PAP written and registered in Aug 2009 before data analysis
- How much detail should be included?
  - Defines exact universe of outcomes, sub-groups and econometric specifications
  - Groups outcomes under hypotheses to facilitate adjustments for multiple inference (i.e. mean effects, standard error corrections)
  - Commits to report all individual outcome results

Omissions and Learning

- We forgot things: added a hypothesis ex post regarding project implementation by drawing together outcomes already in the PAP

- We learned from research fieldwork and piloting: developed new measures of collective action (e.g. SCAs), threw out baseline measures that yielded little variance

- We acquired information about the program from implementation: did not anticipate the focus on skills training, so added new measures to the endline survey

- We added framing to ease interpretation: grouped hypotheses into two intuitive families ex post where we found strong impacts on "hardware" and none on "software"

A Compromise: Limited Flexibility with Full Transparency

- Some flexibility is useful to counter the downside risks of a "purist" approach
  - Rigidity may stifle learning and limits the scope to leverage all available information (including advances in the discipline)
  - Excessive up front costs may deter adoption
- If accompanied by transparency to maintain the credibility of the pre-specification process
  - Reporting of results with and without ex post adjustments
  - Identification of what was pre-specified and when to allow readers to make their own informed judgments
- An observation
  - The CDD study was well powered and had a stark constellation of results, so the compromise approach did not raise awkward questions. It will likely "bite" more strongly in other settings