

Causal Process “Observation”: Oxymoron or (Fine) Old Wine

Nathaniel Beck

*Department of Politics, New York University, 19 West 4th Street, 2nd Floor,
New York, NY 10003*

e-mail: nathaniel.beck@nyu.edu

The issue of how qualitative and quantitative information can be used together is critical. Brady, Collier, and Seawright (BCS) have argued that “causal process observations” can be adjoined to “data set observations.” This implies that qualitative methods can be used to add information to quantitative data sets. In a symposium in *Political Analysis*, I argued that such qualitative information cannot be adjoined in any meaningful way to quantitative data sets. In that symposium, the original authors offered several defenses, but, in the end, BCS can be seen as recommending good, but hopefully standard, research design practices that are normally thought of as central in the *quantitative* arena. It is good that BCS remind us that no amount of fancy statistics can save a bad research design.

1 Introduction

In an important book, Brady and Collier (2004) argue for a unified methodology for both qualitative and quantitative social science that respects both traditions. Obviously, it is hard to disagree with such a laudable goal. However, in a symposium on the book (Beck 2006), I argued that the key contribution of the book, the joining of “causal process observations” (CPOs) to “data set observations” (DSOs) is chimerical. In a rejoinder to my comment, Brady, Collier, and Seawright (2006) argued, using examples from natural science, epidemiology and political science that CPOs can be combined with DSOs in ways that I argued were impossible. Few are interested in the specifics of our disagreement, but many are interested in the vital issues raised by Brady, Collier, and Seawright.¹ I thus rejoin some of the critical issues raised by BCS in the spirit of helping to better understand the relationship of qualitative and quantitative analyses, and where we agree and disagree.

According to BCS, CPOs are the bread and butter of the qualitative analyst; DSOs serve the same role for the quantitative analyst. BCS define a CPO as “[a]n insight or piece of data that provides information about context, process or mechanism, and that contributes

Author’s note: Thanks, for various things, to Henry Brady, David Collier, Gary King, Bernard Manin, Adam Przeworski, and Jas Sekhon. A previous version was presented at the 2006 Annual Meeting of the American Political Science Association, Philadelphia, September 2006. The work was completed while I enjoyed the splendid hospitality of El Centro de Estudios Avanzados en Ciencias Sociales of the Fundación Juan March in Madrid.

¹I will refer both to the original edited volume, the specific contributions of Brady, Collier, and Seawright, and their reply generically as BCS (other than for quotations).

© The Author 2010. Published by Oxford University Press on behalf of the Society for Political Methodology. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

distinctive leverage in causal inference. A causal-process observation sometimes resembles a ‘smoking gun’ that confirms a causal inference in qualitative research, and is frequently viewed as an indispensable supplement to correlation-based inference in quantitative research as well” (Brady and Collier 2004, 227–228). The tying of CPOs to qualitative analysis is strengthened by adding to the definition a reference to “process tracing.”

CPOs are distinguished from DSOs; the latter are the quantitative researcher’s typical measures on a set of variables for each “subject” or “case” in the study. Obviously, quantitative analysts find DSOs to be of value, and qualitative analysts likewise find CPOs to be of value; there is no need to debate this here. Clearly, if we do two separate analyses, and each sheds some light, then the two together must shed more light than either one alone. We do not theorize about political phenomena without observing such phenomena. It is no accident that the greatest advances in the study of legislative bodies were made by students of William Riker and Richard Fenno. The only debate is whether the two types of “observations” can be meaningfully combined in a *single* analysis.

Since it is hard for me to know exactly what a CPO is, I begin by looking at the examples presented by BCS. Although they draw from astronomy and paleontology, epidemiology, and political science, I restrict myself to the most important examples from epidemiology and from the political science examples.

2 Who Gets to Claim John Snow?

In many areas of science, where stochastic variation is irrelevant, scientists proceed by saying “if theory A is correct we must observe X, and so if we do not observe X, theory A must be incorrect.” So errors in the positioning of planets allowed for rejection of Ptolemaic theory (though of course the history of science is a bit more complicated than this). Were BCS simply arguing that we should discriminate between theories by examining their empirical implications, there would be no controversy. I doubt that any reader of *Political Analysis* disagrees with Platt’s (1964) “strong inference” (testing alternative theories by finding situations where they make differing predictions). What this has to do with “CPO” is beyond me.

BCS use the famous work of John Snow on cholera as an example of joining CPOs and DSOs.² Snow is known for his carefully analyzed quasiexperiments showing that cholera was a waterborne illness caused by drinking water that had been contaminated by infected sewage. Snow clearly did make theoretical as well as empirical progress, advancing the contagion theory of cholera to argue that it was spread by something ingested, most likely infected water.

Snow’s belief that cholera was a waterborne illness was based on careful observation, particularly that the disease first affected the alimentary canal and that its rapid spread made simple person-to-person transmission an unlikely route for spreading the infection. We can argue about whether observing that the initial symptoms of a disease are related to

²BCS also discuss the work of Semmelweis. He is most well known for using experimental evidence to show the correct cause of puerperal fever. “Other people before Semmelweis . . . had correctly hypothesized about the nature of puerperal fever . . . Semmelweis is famous amongst epidemiologists then, not so much for the originality of his findings, as for the statistical, scientific approach that he brought to investigating the problem” (Hempel 2006, 169–170). There is nothing in Semmelweis’ methodology that would give pause to the most purely quantitative experimental researcher.

the digestive system bears any relationship to the qualitative analyst’s process tracing, but these “CPOs” were not part of the crucial empirical work that Snow did to *conclusively demonstrate* that cholera was a waterborne disease.

Snow showed that cholera was a waterborne illness in two famous quasiexperiments. In one, he showed that users of a water company that drew from the contaminated portion of the Thames River suffered from cholera at a much higher rate than users of another company that drew from upstream, even though users of the two companies were so mixed that many neighbors used different companies. In the second, he showed that users of the contaminated Broad Street pump were much more likely to get cholera than users of other water sources.³

Snow’s work also shows the danger of drawing conclusions without solid quantitative analysis. David Freedman’s very positive discussion of Snow’s “qualitative” methodology makes this clear. “Of course, he also made his share of mistakes. For example, based on rather flimsy analogies, he concluded that plague and yellow fever were also propagated through water . . .” (Freedman 1991, 299). It was only later that purely quantitative work showed conclusively how yellow fever was actually transmitted and that Snow’s conjecture was incorrect.

Snow’s empirical work was both novel and brilliant but shows “only” excellent quantitative analysis rather than any challenges to such analysis. For example, he could not observe who used the contaminated Broad Street pump, so he plotted cholera outbreaks versus distance to the pump. But rather than using Euclidian distance, he took the distance it would actually take to travel to the pump over available streets; he was thus a precursor of the modern spatial analyst. In his analysis of users of the two water companies, he had to do incredibly hard (and clever) work to figure out which company serviced a given house.

Excellent analyst that he was, he tried to find as many empirical implications of his theory and then see whether the theory was borne out in those cases. For example, he famously found some brewery workers near the pump who did not get cholera since they had no need for any water, and he found some residents of an almshouse who lived under shabby conditions but had access to purer water and also did not get cholera. Testing multiple empirical implications is clearly good but hardly controversial.

The terms “shoe leather” and “detective” are often invoked in retrospections on Snow (Freedman 1991; Hempel 2006). But they are referring to the very hard work involved in careful observation and the collecting of innovative data. There are only DSOs in Snow’s data set. The modern day social science analogue of Snow is the household researcher, carefully gathering multiple types of data on appropriately selected households. Did Snow use some initial observations to see whether his theory or the prevailing “miasma” theory was more consistent with those observations? Of course. But had he stopped there, he would be remembered as one more lucky crank. It is for his empirical work, which is the basis of modern *scientific* epidemiology, that he is remembered. In brief, Snow was

³We need to be careful about CPOs here. Brady, Collier, and Seawright (2006, 364) state that the cholera “epidemic came to a close shortly after Snow convinced the authorities to remove the pump handle, thereby preventing people from using the contaminated water This sequence of specific, localized CPOs, most dramatically the consequences of removing the pump handle, were diagnostic markers that linked the cause (contaminated water) to the effect (cholera).” Why the confirmation of a specific empirical prediction is a CPO escapes me. Thus, it is only amusing to note that Freedman (1991, 295–296) gives a different account. “As the story goes, removing the handle stopped the epidemic and proved Snow’s theory. In fact, he did get the handle removed and the epidemic did stop. However, as he demonstrated with some clarity, the epidemic was stopping anyway, and he attached little weight to the episode.”

a wonderful data collector and quasiexperimentalist, with studies that any quantitative analyst would have been proud to have done. We can now move on to the political science examples.

3 Political Science Examples

Tannenwald, Stokes, and Lieberman

BCS discuss four examples, with Lieberman's (2003) work being added to the three examples discussed in earlier work.⁴ Very briefly, Tannenwald (1999) was interested in explaining why the United States did not use nuclear weapons after World War II. Having four cases and no variation on that dependent variable, she turned to documents to see what accounts policy makers gave for why they did not use nuclear weapons in a crisis. Reasonable people can differ on the utility of accounts given by decision makers as to why they did what they did; sometimes they tell stories we like, and we are happy, and sometimes not. So a study of what policy makers said about why they did not want to use nuclear weapons is clearly interesting, but it is a different study from (the impossible one of) the causes of the United States using or not using nuclear weapons after World War II.

Similarly, Stokes (2001) had one quantitative analysis of elections and economic policy in Latin America, and one qualitative analysis on the account that some leaders gave about why they pursued the policies they did. Both useful, but it makes no sense to talk of Stokes *adjoining* qualitative data to her quantitative study. In studying the role of electoral politics in Federal reserve policy making, how much interest should we have in finding or not finding the "smoking gun" where Arthur Burns either did or did not tell his colleagues that Nixon cared a great deal about his own reelection; does the presence or absence of a transcript here matter? Is not the quantitative evidence about the timing of observable policy actions not the relevant evidence here (Beck 1987)?

Although we are all happy when political actors give an account which is consistent with our theories, the value of these accounts is somewhat unclear. To take a simpler situation, imagine a study of why members of Congress facing serious scandals choose not to run for reelection. Imagine previous decades where there are very few such scandals, and the three or four scandalized all chose not to seek reelection. Would we be as happy with the conclusion that members of Congress facing a scandal chose to spend more time with their families as with the CPOs of Stokes or Tannenwald? If there were enough retirements for a standard quantitative analysis, would we wish to adjoin to it these CPOs?

The newly discussed study of Lieberman is a comparative analysis of the tax capacity of various nations. He began with a standard cross-country regression and found that tax revenues were well predicted by economic development. He then looked closely at two cases, South Africa and Brazil, with similar levels of economic development but very different tax revenues. He then hypothesized that elites are willing to pay taxes if they believe that they will receive a nontrivial portion of the subsequent government spending, which he then tested with another regression. The moves between case study and regression are not uncommon. But to say that Lieberman adjoined CPOs to DSOs seems to be stretching the idea of adjoining. Obviously, Lieberman devotes relatively more effort to case study than do

⁴As in my earlier comments, my issues here are purely methodological and I have no interest in critiquing any of these works, which were chosen because BCS discussed them.

most quantitative political scientists, but there is nothing in his work to challenge the standard ideas of inference.

Brady versus Lott

Finally, let us turn to Brady's reanalysis of Lott's finding on the impact of the early election call in Florida in 2000.⁵ In my earlier comment, I noted that there was nothing in Brady's discussion that was not in the standard quantitative analysts' toolkit. We all believe that a call of an election outcome at 8:45 p.m. cannot change the turnout of those who voted before that time. Brady combined that, with standard quantitative studies of media and turnout (Jackson 1983, e.g.), to conclude that Lott drastically overstated the maximal possible impact of the early call. Unless the standard quantitative toolkit is not allowed to require temporal precedence for causality and is not allowed to note that a necessary condition for media to have an effect is that people receive the message, I simply do not see where the Brady's reanalysis draws on adjoining CPOs to DSOs.

The Florida case is an interesting one for students of media effects. It might be amenable to standard survey analysis similar to Jackson's study of the 1980 early concession of Jimmy Carter (before the California polls closed). Brady's calculations indicate that such a survey would likely be infeasible; the calculation of a minimum sample size necessary to find a statistically significant effect given a plausible substantive effect is a routine quantitative tool.

One could also have (easily) improved on Lott's difference-in-difference design. Analysts using difference-in-difference designs work hard to find units that are similar before an intervention. My intuition about Florida is that few would have been so brave as to assume that Dade and Okaloosa Counties could be compared. Only a few more would be brave enough to compare them after a regression adjustment. Perhaps some attempt at matching might improve matters, but it would take a lot of optimism to believe that one could create a synthetic Okaloosa County from counties in the Eastern Time Zone.

Snow was lucky in having two water companies with different intake points but serving the same streets. Here, an analyst might be lucky enough to find some election districts, which appear similar but are in different time zones. Sometimes the gods of geography favor quasiexperimenters, sometimes they do not.

But the critical point is that we can all agree that Lott's research design was poor. Difference-in-difference designs may well be a staple of modern quantitative research, but not any such design is a good one. A staple of modern quantitative research is both examining and finding tools to examine when difference-in-difference designs are adequate, or, more importantly improving on such designs (Abadie, Diamond, and Hainmueller 2010, e.g.). We need not disagree about the quality of Lott's difference-in-difference design to disagree about whether CPOs should be adjoined to Lott's study.⁶

⁵It is a bit embarrassing to spend so much time discussing Lott's analysis, which appeared in an Op-Ed in the Philadelphia Inquirer, which is not, to the best of my knowledge, peer reviewed. Brady clearly bested Lott; this was not hard.

⁶Brady suggests that these CPOs would come from interviewing election officials or studying media reports. Unless such people have a magical machine, which can create analyses that researchers cannot create, what could we learn from such a study, other than what the media or election officials think is true? As with Tanenwald, perhaps an interesting question, but rather far from the one originally posed.

4 Conclusions

So where does this leave our discussion? Oddly, we all agree that Snow did wonderful research and are simply arguing as to whether it was quantitative or qualitative. If we were to drop the terms qualitative and quantitative and simply say that everyone should strive to do research like Snow's, I would be perfectly happy.

Similarly, we all agree that mindless regression running is not a good way to proceed. Theory (whether high formal or low intuition) is needed to know what variables go into a regression. There is a lot of art in figuring out what is the right regression design, and little of that art is really statistical.

Many of the examples of BCS argue that an experimental or quasiexperimental approach is superior to a large N regression-based approach. This does not challenge the thinking of standard quantitative research. It is good that political methodologists now worry more about good research design than the latest small improvement in a complicated statistical model. So if this is what BCS are arguing for, great. If BCS are simply recreating Platt's half a century old (but of course still valid) call for "strong inference," who could disagree? A more modern statement of the same idea is Angrist and Pischke's (2009) call that observational research should always start with a researcher asking what an experiment, if feasible, would look like. Research design is clearly fundamental, and no amount of fancy statistics can save a bad research design.

No one could argue that knowledge of how the world works is a bad thing, and this knowledge is at least partially acquired through observation of real political phenomena. It was not bad that Snow intuited that a disease that attacked the intestines was likely to be caused by something ingested; of course, it was better when biologists gave this a firm theoretical foundation years after Snow's work. So models of turnout should include mobilization efforts and models of taxation should include more than per capita gross domestic product.

My only disagreement with BCS is I still have no idea what to do with CPOs in a *quantitative analysis*, that is, how can they be adjoined in a meaningful way to DSOs. I am happy to agree that, when possible, experiments and quasiexperiments are often (usually, almost always) superior designs to large N regressions. We should hold up Snow as an example to all of our students. Good research design is fundamental. What this has to do with "qualitative" methods is beyond me. But other than nomenclature, we seem to agree on what good empirical research looks like.

References

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program. *Journal of the American Statistical Association* 105:493–505.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics*. Princeton, NJ: Princeton University Press.
- Beck, Nathaniel. 1987. Elections and the fed: Is there a political monetary cycle? *American Journal of Political Science* 31:194–216.
- Beck, Nathaniel. 2006. Is causal-process observation an oxymoron? *Political Analysis* 14:347–52.
- Brady, Henry E., and David C. Collier, eds. 2004. *Rethinking social inquiry: Diverse tools, shared standards*. Lanham, MD: Rowman and Littlefield.
- Brady, Henry E., David Collier, and Jason Seawright. 2006. Toward a pluralistic vision of methodology. *Political Analysis* 14:353–68.
- Freedman, David A. 1991. Statistical models and shoe leather. *Sociological Methods* 21:291–313.
- Hempel, Sandra. 2006. *Medical detective: John Snow and the mystery of cholera*. London: Granta.
- Jackson, John. 1983. Election night reporting and voter turnout. *American Journal of Political Science* 27:615–35.

- Lieberman, Evan S. 2003. *Race and regionalism in the politics of taxation in Brazil and South Africa*. New York: Cambridge University Press.
- Platt, John R. 1964. Strong inference. *Science* 146:347–53.
- Stokes, Susan Carol 2001. *Mandates and democracy: Neoliberalism by surprise in Latin America*. New York: Cambridge University Press.
- Tannenwald, Nina. 1999. The nuclear taboo: The United States and the normative basis of nuclear non-use. *International Organization* 53:433–68.