The Design-Based Revolution in Comparative Politics

Paul Kenny

Chapter forthcoming in N. Farrelly, A. King, M. Wesley, and H. White (eds.) *Muddy Boots and Smart Suits: Researching Asia-Pacific Affairs*, Singapore: Institute of Southeast Asian Studies

Introduction

Is aerial bombardment an effective counterinsurgency tactic? Do immigrants depress local wages? Does the presence of natural resources harm democratic consolidation? Answering questions such as these ones requires uncovering general patterns of cause and effect. Social scientists have been attempting to put this search for causal inference on a sure philosophical and methodological footing at least since Durkheim's pioneering research on suicide in the nineteenth century, but progress has been slow, coming in fits and starts. Even the diffusion of computing technology and the modelling revolution it brought about did not solve the problem. However, the last decade or so has seen one of the most exciting developments in the history of causal inference in the social sciences for generations.

The design-based revolution has transformed how social scientists go about their research and the kinds of causal inferences they can now draw from it. Field experiments and natural experiments are at the foundation of design-based inference. A growing body of scholarship outlining the principles, techniques, and promise of this kind of research already exists, but most of it has been aimed at an audience already well-versed in comparative analysis and statistical methods (Gerber and Green 2012, Morgan and Winship 2014, Angrist and Pischke 2015, Imbens and Rubin 2015). The logic of design-based inference is actually very intuitive, despite its somewhat daunting technical language. In this chapter I want to introduce the design-based approach to a non-specialist audience, primarily those who use qualitative rather than quantitative methods.

I first introduce the two main models of causal inference common to comparative politics, the sufficient-component-cause (SSC) model and the counterfactual model. Second, I outline the problem of confounding in traditional qualitative and quantitative methods in the social sciences. Third I detail the experimental method in its basic form and present some examples of randomized controlled trials (RCTs) in Asia. Fourth, I discuss extensions to the experimental model in the form of natural experiments and again illustrate the method with some examples from Asia. Finally, I discuss the limitations of the design-based revolution and what needs to be addressed going forward.

Backward and Forward Inference

There are two main models of causal inference used in comparative politics, the SSC model and the counterfactual model. These models are sometimes associated with qualitative and quantitative methods respectively, but this understanding is actually incorrect. The difference between the approaches is better thought of in terms whether the investigator reasons backwards or forwards.

In backward reasoning, we first observe an outcome and then attempt to uncover the processes or conditions that led to that outcome. This approach makes sense for discrete historical problems. Like a detective investigating a crime, the researcher sifts through the available evidence to trace through a sequence of events. This would be the approach used to explain why a particular counterinsurgency campaign failed or why a particular coup occurred. In some cases, this can lead to persuasive and reliable historical inference. However, even though this approach is utilized to answer distinct historical questions, it cannot work without more generalizable models of causal inference. That is, how we infer causality in a particular chain of events is typically shaped by how we believe people should behave on average in a given situation. Thus, even historians cannot fully escape the rules of good causal inference.

The more general form of the backward looking approach is sometimes called the SSC model. This model of causal inference is based on the search for causes of effects. That is, we know the effect or outcome, and want to elucidate its cause(s). For this approach to work, we need more than one case, so that we can identify sufficient and necessary conditions that might produce the outcome in question.

A sufficient condition is a condition such that its occurrence is always enough by itself to produce the outcome in question. If A is the condition and B the outcome, we can say, *if* A *then* B. Conditions may be jointly sufficient if the presence of two or more conditions always produces the outcome in question. If C is another condition, then *if* A *and* C *then* B. A necessary condition is a condition such that an outcome will never occur unless this condition is present. Simply, B *only if* A. Conditions may also be jointly necessary. B *only if* A *and* C. This approach is thus attractive when researchers are looking to explain the full configuration of factors that caused particular events or outcomes.

In qualitative research, this approach works by applying either the method of difference or the method of agreement. The method of difference works by observing the occurrence of an outcome in one case, and its non-occurrence in another number of cases, where the cases have all conditions in common except one in the former; this one differing condition is the cause of the differing outcome. The idea, in short, is to match treatment and control groups along every dimension except that the former experience the treatment while the latter do not. The approach attempts to mimic the modelling techniques of regression analysis but on a smaller scale (King, Keohane et al. 1994). The method of agreement instead works by observing the occurrence of the same outcome in two or more cases, where the cases have no conditions in common except one; this one similar condition is the cause of the outcome.

An example of this approach is Dan Slater's *Ordering Power.* Slater looks to identify a set of conditions that explain why the authoritarian regimes of Southeast Asia varied in their stability over the late twentieth century (Slater 2010). Slater identifies defensive elite protection pacts in response to the revolutionary conditions of the mid-twentieth century as the key condition that separated the authoritarian leviathans of Singapore and Malaysia from the less stable authoritarian regimes of the Philippines and Indonesia.

Problems arise when attempting to formulate a generalizable causal model from one or a few discrete cases. Many outcomes of interest in political research are highly complex, and there is rarely a condition, or set of conditions, that is sufficient to produce them. Moreover, such conditions are especially difficult to identify when working with a small number of cases (King, Keohane et al. 1994). However, the search for necessary and sufficient conditions is not the exclusive preserve of qualitative research. Indeed, the SSC model was formally elucidated for research in epidemiology that typically works with much larger datasets (Rothman 1976). The difficulties of inferring causality with this model of inference are thus common to both qualitative and quantitative research. Although a great deal of political and sociological research takes this approach, and while it is consistent from an epistemological and ontological perspective, in practice it is difficult if not impossible to implement without introducing several sources of bias. Bias in this sense does not refer to the prejudices of the researcher. Rather, bias refers to the tendency to over- or under-estimate the value of a causal factor because of the way in which the data has been processed. I discuss sources of bias in more detail below.

While it is enticing to search for causes of effects, often this research is often most valuable when it suggests avenues for new kinds of forward inference. In this chapter I am primarily concerned with the latter. The other way of reasoning is to take a particular cause and then measure its effects. Causal inference in this model is established when we can compare two states of the world, a treated state, and a non-treated, or control state. The difference between the two is called the treatment effect. Of course, we cannot observe both states of the world at the same time, so we have to compare the observed outcome with a potential outcome. This is the counterfactual model of causal inference and it is at the heart of the design-based revolution that I discuss further below.

Counterfactual models of causation underpin a lot of research in both quantitative and qualitative political science. The great difficulty in drawing causal inference from observations of the real world is the presence of confounding factors. This is just as true of forward-looking causal models as it is of backward-looking models. Design-based inference attempts to implement or replicate as closely as possible the experimental method in order to eliminate this problem of confounders.

Confounding in Comparative Politics

Confounding occurs when the causal factor in question is not independent of other potential causal factors. If the cause, or treatment, is not independent (or exogenous) from other factors that might have led to the outcome in question, then it isn't possible to accurately infer the effect of that cause. The main sources of confounding that arise in comparative politics are selection bias and endogeneity bias.

Selection bias refers to the selection of individuals, states, or other units of analysis such that the sample used is not representative of the population under investigation (Geddes 1990). If we were attempting to study the effectiveness of a particular drug, and deliberately selected only those subjects who evidenced a response to it for analysis, the results would be biased. Similarly, in social science the arbitrary restriction of the sample of units of analysis, say by geography or by time period, can have a distorting effect on the results of an empirical study.

Take Maya Tudor's recent book on the divergent post-independence regime trajectories of India and Pakistan, *The Promise of Power* (Tudor 2013). While India has been mostly democratic, Pakistan has been highly unstable, reverting from democracy to authoritarianism and back again. For Tudor, the explanation centers on the political movements that led India and Pakistan's respective independence movements, specifically, the relative strengths of the party networks of the Indian Congress Party on the one hand and the Muslim League on the other, and the class basis of the two respective political movements. However, with such a truncated sample (of two), it is impossible to determine whether this explanation is correct. Although India and Pakistan are matched in many respects, there are many sources of underlying variation between them such that all possible counfounders (i.e. additional variables that are correlated with both cause and outcome) cannot be controlled for. Indeed, looking at the same two cases, Philip Oldenburg (2010) comes up with a very different account. Oldenburg also goes back to the contingent historical process that resulted in the greater relative strength of the Congress network in comparison to that of the League, but he is primarily interested in explaining the relative strength of the Pakistani bureaucracy when compared to India. He argues that those areas of colonial India that became Pakistan were governed through a more bureaucratically dominant administrative structure than the areas that became India. If we were to look at more cases, even within British colonial Asia, such as Malaysia, Myanmar, and Sri Lanka, different conclusions would likely emerge. Expanding the sample to all of the British Empire, or to all of the colonies of all empires, would alter the results even further. Even then, however, we could not be sure that there were not further omitted variables that were correlated with both the nature of colonial independence movements and the nature of postcolonial regime trajectories.[1] This leads us to the related but distinct problem of endogeneity.

Endogeneity refers to the problem of including a causal factor in the model that is potentially correlated with other unobservable causal factors relegated to the error term. In other words, it means that some potentially important causal factor that influences both cause and effect is excluded from the model, thus misrepresenting the true effects of those factors included in the model.

Omitted variable bias is a common source of endogeneity. In the example of attempting to study the effectiveness of a drug, it could arise from allowing subjects in a trial to choose whether take the drug or the placebo. Those individuals willing to risk taking a new drug might be sicker, thus influencing how effective the drug would appear to be.

[1] For more on selection bias, see the chapter in this volume by Charles Miller.

To illustrate its operation in comparative politics, it is easiest to take an example from the literature. Say we want to know the effect of corruption on economic development. In *Crony Capitalism*, David Kang (2002) asks why corruption is sometimes associated with high growth and sometimes with low growth. Comparing South Korea and the Philippines, Kang argues that different types of corruption have different effects on development. The presence of a small and stable network of business and government elites in South Korea meant that corruption functioned to lower transaction costs and promote growth. A dispersed set of conflicting elites and bureaucrats in the Philippines meant that corruption raised transaction costs and inhibited growth. Kang's causal model is certainly plausible. But what if there was some prior condition that simultaneously caused South Korea to have a particular type of business and government elite structure *and* to be more developed, but which caused the Philippines to be less developed? Recent research by Jong-sung You (2015) suggests that the greater equality of landholding in South Korea might have done exactly that. Thus, without taking account of this omitted variable, we would wrongly infer that the type of business and government elite structure caused development when in fact this was an intervening factor at best or epiphenomenal at worst.

Endogneity bias also occurs when the outcome of interest recursively affects the cause of interest. That is, there is a feedback between cause and effect. In her recent historical investigation of the Pakistan military, Christina Fair argues that the strategic culture of Pakistan's army accounts for its persistently belligerent behaviour in the region, even though its actions have brought it limited success (Fair 2014). Although Fair is incredibly thorough in her investigation of the relevant sources, there remains the strong possibility that the army's strategic culture, or set of beliefs, is affected over time by the very outcomes Fair is attempting to explain. That is, the relationship between the Army's strategic culture and the outcomes of its military and strategic engagements is circular. This is a widespread problem with observational research, especially research in which the effect of beliefs on behaviour is concerned.

It is important to note that these kinds of confounding can occur in quantitative as well as qualitative research. In quantitative research, investigators attempt to control for potential confounding variables by including a large number of cases so that only like cases are compared. This means that cases are effectively matched on all of those factors that the researcher believes might influence the result. Ultimately, however, the addition of more cases does not resolve the issue, nor does the application of more and more sophisticated modelling techniques. In fact, it turns out that no amount of modelling can really overcome them because the researcher cannot be sure that there isn't some unobserved confounder that is influencing both cause and effect. How then can we get around this problem of confounding?

The Experimental Method

The simplest way to overcome the problem of confounding is through randomization. The random assignment of units of analysis to treatment and control groups ensures that the treatment is uncorrelated with possible confounding variables, as the latter average out across the two groups. This implies the use of the experimental method. The experimental method provides very reliable causal inference. Because the treatment is directly manipulated by the researcher, she can be confident that it is unrelated to potential confounding variables. For instance, in a drug trial, it means that individual participants are given no choice as to whether they will take the treatment or the placebo. Any potentially important extraneous factors such as genetic variation or underlying sickness should therefore be the same on average across treatment and control groups. The subsequent variation in outcome, if there is one, can thus be attributed to the effect of the drug.

The experimental method typically eliminates confounding by having large enough study samples such that treatment and control groups resemble one another on average. With a very small sample, the researcher could inadvertently assign the more healthy individuals to take the drug rather than the placebo. Most experiments in the political sphere are therefore large-N (number of units of analysis) rather than small-N studies. While they are thus typically quantitative studies, the actual number of units of analysis doesn't need to be exceedingly large. Moreover, as I discuss below, the principles of counterfactual causal inference translate over into small-N qualitative research.

Randomized Control Trials

In the real world, the experimental method means applying an intervention in the form of a field experiment or randomized control trial (RCT) (Gerber and Green 2012). Some of the earliest systematic research employing RCTs in political science studied the effect of various kinds of voter mobilization techniques on voter turnout (John 2013). Interventions that involve the provision of information tend to be more common as they are cheaper to implement than other interventions. As the effectiveness of RCTs has come to be appreciated in the policy world, however, recent years have seen the deployment of a large number of ambitious and exciting projects.

An RCT aims to make an inference about this causal effect in a population based on the results from a sample of that population. The first task in an RCT is to select a sample from the population. This is not unproblematic. The sample should resemble the population for whom the treatment is relevant in important dimensions like age, gender, and so on. To ensure this representativeness, there should be randomization in the selection of trial participants. Because of the law of large numbers, randomization ensures that the sample average should approximate the population average. The second step is to select treatment and control groups. Again, randomization is critical. We cannot allow participants to self-select into the treatment. Treatment and control groups should then be checked for balance to make sure that we have not, by some fluke, assigned only people of one ethnicity or gender to the treatment, and only people of another ethnicity or gender to the control. This process ensures that there are no confounding factors that might bias our results. The third step is to measure the average outcome in the control group and the treatment group. The difference in these averages is called the average treatment effect.

Some Examples

In one of the best-known examples of an RCT in the political economy field, Benjamin Olken investigated how corruption might be reduced in the infrastructure sector. Olken (2007) randomly assigned 600 Indonesian village road projects to control and various treatment groups. Treatments were of two types: greater top-down auditing of project expenditures and construction, and civil society monitoring of the projects. Corruption was estimated by the difference between actual expenditures on the village road projects and the actual cost of construction based on an experimental road constructed explicitly as a control for the project. Olken found that while top-down monitoring reduced leakage through corruption as compared to the control group, grassroots monitoring was only effective in very limited circumstances.

In another highly ambitious project, Banerjee, Duflo et al. (2015) investigate the effects of a comprehensive developmental package for the extreme poor on their subsequent welfare. The package included a productive asset grant, training and support, life skills coaching, temporary cash consumption support, and typically access to savings accounts and health information or services. The study included on 10,495 villagers in six developing countries including India and Pakistan. The researchers measured the impact of this package on consumption, food security, productive and household assets, financial inclusion, time use, income and revenues, physical health, mental health, political involvement, and women's empowerment. The study found statistically significant impacts on all 10 key outcomes that persisted in most cases a full three years after the initial intervention. Although the study was expensive to implement, they found that in most countries the extra earnings of participants outweighed the program cost. This suggests that well-targeted poverty alleviation programs can actually be very effective.

Critiques of RCTs

RCTs clearly contribute positively to our knowledge of cause and effect in many areas of policy and politics. However, they are not without substantial problems. First, RCTs in the policy space come with major ethical implications (Barrett and Carter 2014). RCTs could have potentially adverse consequences, despite the best intentions of ethics review boards. Second, while RCTs are solid on internal validity, they are necessarily inhibited by a lack of external validity (Garcia and Wantchekon 2010). That is, even if a policy shows a positive effect in one trial in one location, we cannot be sure it will work in another time and place until further experiments are carried out (Cartwright and Hardie 2012). Third, in the real world, information and resources for individuals fit within a much broader sociological context, many RCTs find only substantively weak effects. Their impact is deeply contingent on a wide set of factors that are not easily controlled for, even in RCTs. Fourth, interventions can only be contemplated in a restricted set of areas of politics and economics (Rodrik 2008). We cannot use an RCT to study the effect of a war or a new electoral system. Fifth, and more fundamentally, because these are real world interventions, we would often like to understand the mechanism that is generating the outcome (Deaton 2009). To say that the discovery of a new education policy that increases high school completion would be incredibly welcome is an understatement. But we would need to know how it works. What if it increased high school completion rates at the expense of individual health because it was based on the universal prescription of Adderall or some other attention-increasing drug? In short, RCTs on their own are no panacea for causal inference without a good understanding of causal mechanisms.

Natural experiments

Natural experiments are closely related to field experiments in the way in which they attempt to draw causal inference (Dunning 2012). Although the interventions are not controlled by the researcher, they are selected because they are implemented in an as-if random way. Because such natural experiments are contextually embedded, they offer a chance to see how large-scale projects work in practice. Natural experiments can draw on a variety of sources. Some of them rely on policies delivered by governments in random or quasi-random ways. Others instead draw on exogenous shocks like natural disasters. The key distinction between a natural experiment and a purely observational study is that the treatment (or cause) is uncorrelated with confounding variables that may also affect the outcome.

As with RCTs there are many good examples. In India, in order to increase women's participation in politics, a 1993 law reserved leadership positions on randomly selected local village councils for women. Using a survey of 8,453 of adolescents aged 11-15 and their parents in 495 villages, Beaman, Duflo et al. (2012) find that compared to villages that never had a reserved female leader, the gender gap in aspirations closed by 25 percent in parents and 32 percent in adolescents in villages assigned to a female leader for two election cycles. Female adolescents in treated villages spent less time on household chores than in non-treated villages, although there was no sustained impact on women's labor market opportunities in reserved villages. The effect of reserving village council leadership positions seems to have persistent political effects also. Bhavnani (2009) finds that the probability of a woman winning office in previously reserved councils is approximately five times greater than the probability of a woman winning office if the constituency had not been reserved for women.

Extensions of the natural experimental approach include regression discontinuity designs and instrumental variables (IV) designs. Regression discontinuity designs rely on exploiting another kind of as-if random assignment to treatment and control groups. A number of studies have attempted to determine the impact of migration on the prosperity of migrants and their families to whom they remit earnings. An obvious problem with many studies would be that the families of those who migrate could be systematically different from those who do not. Michel Clemens and Erwin Tiongson (2014) study Filipinos migrating to work in manufacturing jobs in South Korea through a bilateral agreement under the Employment Permit Scheme (EPS) program. Because the EPS requires exceeding a certain minimum score on a language competency exam, it is possible to compare the effect of migration on those individuals who just exceed the score to those who just fall short. Because the difference in score is so similar, we can assume that the individuals are on average similar on things that matter like ambition and intelligence. They find that migration has important effects on households' financial behaviour, including tripling expenditure on education and health, reducing borrowing, and raising savings. Interestingly, the shifting of budget allocation decisions from husbands to wives appears to drive much of this change, independent of the effect of remittances.

IV analyses also attempt to overcome the problem of confounding by exploiting exogenous sources of variation in causal forces. However, in this case, the actual causal variables of interest are not randomly assigned. Rather, another variable or set of variables that is related to the causal variable of interest but *not* directly to the outcome of interest, is randomly assigned. It is common in such research designs to use factors like weather or other natural events. For example, Yusaku Horiuchi and Jun Saito (2009) revisit the question of whether voter turnout affects election and policy outcomes. A typical observational study would suffer from confounding because of omitted variable bias and endogeneity bias due to the repeated interactions between politicians and voters. Rainfall is known to suppress voter turnout but it is obviously unaffected by policy outcomes. Rainfall thus functions as an instrument for turnout. Using a large, municipality-level dataset from Japan, Horiuchi and Saito show that the effect of turnout on intergovernmental fiscal transfers is large and statistically significant.

Although in each of these examples, the researcher has not been able to directly manipulate assignment to treatment and control groups, because the assignment process was not influenced by the characteristics of the subjects in question, we can be quite confident that any variation potentially important additional factors average out between the two groups. Thus, the observed differences in outcome are due to the effect of the treatment.

Designing Comparative Research

Like field experiments, natural experiments draw causal inferences by estimating the average treatment effect across a large number of units of analysis (whether individuals, families, districts, nations, or something else). They thus have a natural affinity with quantitative approaches. Where the experimental design is particularly clean, the statistical analysis might not be much more complicated than a simple difference-of-means test between treatment and control groups. While the counterfactual approach to causality works especially well with quantitative studies, it also has application in qualitative research. In fact, the principles of the design-based revolution suggest that researchers could instead look to identify cases in which causal forces of interest are exogenously assigned.

In fact, there are innumerable historical shocks and policy interventions that can be studied in this way from a qualitative approach. Natural disasters, commodity price shocks, and mass displacements of citizens in neighbouring countries are examples of exogenous shocks that can be used to study to the effects of causes in qualitative research. In my own research, for instance, I have studied the impact of an institutional change in center-periphery relations in India precipitated by the death of Prime Minister Jawaharlal Nehru (Kenny 2013). Based on an extensive reading of the primary and secondary source material, I argue that Nehru's death acted like an exogenous shock, shifting the power balance away from the Union government and towards the state governments. Similarly, in another ongoing study, I examine the effect of an Indonesian Supreme Court decision to open party lists on the power of national parties over their subnational branches.

Such research might entail comparing across units of analysis. For instance, the sudden discovery of oil in one region of a country versus its non-discovery in adjacent regions would allow researchers to explore the effects of natural resources on the local economy and society in a way not possible with simple cross case comparative research. Inter-temporal comparisons can also play an important role in causal inference. The exogeneity of the treatment is central to causal inference. For instance, economists have been investigating the effects of the 2011 earthquake on the Japanese economy using forecasting methods, but this work could also be done from a qualitative perspective.

This kind of qualitative counterfactual analysis should not be seen as the lesser cousin of quantitative counterfactual analysis. As critics of the experimental method argue, simply establishing a correlation between a treatment and outcome variable should not be the end of the investigation. We need to know how and why such correlations might exist which demands more detailed probing of the causal mechanisms at work. Qualitative research is especially suited to this kind of investigation, whether in the form of in depth comparative ethnographic research or detailed archival research.

Conclusion: The Promise and Limits of the Design-Based Revolution

It would be rash to completely dismiss all research not based on an experimental or quasi-experimental research design. We can learn a lot from observational study. Indeed, experiments are often only suggested by careful attention to causal processes in the real world. However, the design-based revolution cautions all scholars to be much more attentive to the issues of confounding. A well-designed study does far more to mitigate (though not eliminate) the problem of confounding in quantitative research than any number of modelling techniques. Case selection in qualitative research should also be influenced by this discovery. Exogenous shocks can play a useful role in identification of causal factors in qualitative as well as quantitative research. The differential impact of a policy intervention, an external political event, or a natural disaster on otherwise similar political units can similarly inform causal reasoning in qualitative studies. It suggests a slightly different logic of case selection than one based on the method of difference alone.

The design-based revolution thus promises to continue to improve causal inference across the social sciences, but it is not without its potential drawbacks. There is a risk that researchers will tend to favour questions for which natural experiments exist or topics in which field experiments can be carried out. Many problems central to the political world could be ignored if design was to become the only criterion for selecting a subject of study. Much of what happens in the social world is complex. By this I don’t mean to say that it is complicated. That is a different matter. Complexity refers to the fact that what we call cause and effect are often not so neatly separable. In the real world, policy interventions are filtered through a maze of institutions, belief systems, and private interests that make their effects difficult to predict. Critically, under conditions of complexity, the totality of an outcome is not simply a result of the sum of its parts. Events, conditions, and processes interact. Thus, as in the disciplines of ecology and climatology, complex systems can only be partially studied through experimentation in which single elements of a system are manipulated. In many cases, whole systems have to studied together, often relying on the use of observational data. Observational and experimental studies should be seen as complimentary in developing an understanding of cause and effect in the social and political world.

References

Angrist, J. D. and J. r.-S. Pischke (2015). Mastering 'metrics : the path from cause to effect. Princeton, Princeton University Press.
Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert and C. Udry (2015). "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." Science **348**(6236): 1260799.
Barrett, C. B. and M. R. Carter (2014). Retreat from Radical Skepticism: Rebalancing Theory, Observational Data and Randomization in Development Economics. Field experiments and their critics : essays on the uses and abuses of experimentation in the social sciences. D. L. Teele. New Haven, Yale University Press: viii, 270 pages.
Beaman, L., E. Duflo, R. Pande and P. Topalova (2012). "Female leadership raises aspirations and educational attainment for girls: A policy experiment in India." science **335**(6068): 582-586.
Bhavnani, R. R. (2009). "Do electoral quotas work after they are withdrawn? Evidence from a natural experiment in India." American Political Science Review **103**(01): 23-35.
Cartwright, N. and J. Hardie (2012). Evidence-based policy : a practical guide to doing it better. Oxford ; New York, Oxford University Press.
Clemens, M. and E. Tiongson (2014). Split Decisions: Household Finance when a Policy Discontinuity Allocates Overseas Work, Center for Global Development.
Deaton, A. S. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development, National Bureau of Economic Research.
Dunning, T. (2012). Natural experiments in the social sciences : a design-based approach. Cambridge ; New York, Cambridge University Press.

Fair, C. (2014). Fighting to the End: The Pakistan Army's Way of War. New York, Oxford University Press.
Garcia, F. M. and L. Wantchekon (2010). "Theory, external validity, and experimental inference: some conjectures." The ANNALS of the American Academy of Political and Social Science **628**(1): 132-147.
Geddes, B. (1990). "How the cases you choose affect the answers you get: Selection bias in comparative politics." Political Analysis **2**(1): 131-150.
Gerber, A. S. and D. P. Green (2012). Field experiments : design, analysis, and interpretation. New York, W. W. Norton.
Horiuchi, Y. and J. Saito (2009). Rain, Elections and Money: The Impact of Voter Turnout on DIstributive Policy Outcomes in Japan. Asia Pacific Economic Paper, Australian National University.
Imbens, G. and D. B. Rubin (2015). Causal inference for statistics, social, and biomedical sciences : an introduction. New York, Cambridge University Press.
John, P. (2013). Field Experiments in Political Science Research. Afailable at SSRN http://dx.doi.org/10.2139/ssrn.2207877.
Kang, D. C. (2002). Crony capitalism : corruption and development in South Korea and the Philippines. Cambridge ; New York, Cambridge University Press.
Kenny, P. D. (2013). The Patronage Network: Broker Power, Populism, and Democracy. PhD PhD Dissertation, Yale University.
King, G., R. O. Keohane and S. Verba (1994). Designing social inquiry : scientific inference in qualitative research. Princeton, N.J., Princeton University Press.
Morgan, S. L. and C. Winship (2014). Counterfactuals and causal inference : methods and principles for social research. New York, Cambridge University Press.
Oldenburg, P. (2010). India, Pakistan, and democracy : solving the puzzle of divergent paths. New York, Routledge.
Olken, B. A. (2007). "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." Journal of Political Economy **115**(2).
Rodrik, D. (2008). The New Development Economics: We Shall Experiment, but how shall we learn? unpublished manuscript, Kennedy School, Harvard University.
Rothman, K. J. (1976). "Causes." American Journal of Epidemiology **104**(6): 587-592.
Slater, D. (2010). Ordering power: contentious politics and authoritarian leviathans in Southeast Asia. New York, Cambridge University Press.
Tudor, M. J. (2013). The promise of power: the origins of democracy in India and autocracy in Pakistan. New York, Cambridge University Press.
You, J.-s. (2015). Democracy, inequality and corruption : Korea, Taiwan and the Philippines compared. New York, Cambridge University Press.