

## case selection and the comparative method: introducing the case selector

*timothy prescott and brian r. urlacher\**

Political Science Department, University of North Dakota, 293 Centennial Drive,  
Grand Forks, ND, USA

E-mails: [timothy.prescott.und@gmail.com](mailto:timothy.prescott.und@gmail.com); [brian.urlacher@und.edu](mailto:brian.urlacher@und.edu)

\*Corresponding author.

<https://doi.org/10.1057/s41304-017-0128-5>; published online 14 August 2017

### Abstract

We introduce a web application, the Case Selector (<http://und.edu/faculty/brian.urlacher>), that facilitates comparative case study research designs by creating an exhaustive comparison of cases from a dataset on the dependent, independent, and control variables specified by the user. This application was created to aid in systematic and transparent case selection so that researchers can better address the charge that cases are 'cherry picked.' An examination of case selection in a prominent study of rebel behaviour in civil war is then used to illustrate different applications of the Case Selector.

**Keywords** comparative method; case selection; software; qualitative methods

In his seminal article on the comparative method, Arend Lijphart (1971) identifies and discusses four challenges in the application of the comparative method to the study of politics. First, he critiques the discipline for limited methodological awareness. Second, he points out that it is difficult to identify cases that are perfectly similar or dissimilar, which makes it problematic to apply Mill's logic of difference and logic of concurrence.<sup>1</sup> Third, he stresses that the

nature of causality in the social world is probabilistic, so negative findings do not provide sufficient reason to reject a hypothesis. And fourth, Lijphart wrestles with how to handle the flood of cases and data that a social scientist must navigate in the selection of cases.

The first of these problems is far less of a concern today. A robust conversation about methodology has been at the heart of the discipline for more than two decades, and tremendous progress has

been made in qualitative, quantitative, and formal methodologies.<sup>2</sup> While the push toward more sophisticated qualitative research designs has somewhat displaced the comparative method (see Brady and Collier, 2004), there is also a recognition of the value of the comparative method in mixed method research designs (see Slater and Ziblatt, 2013; Tarrow, 2010).

The second of the four problems identified by Lijphart is simply intractable. The world is what it is, and the dogged social scientist must make do. To some extent this is also true of the third concern raised by Lijphart regarding probabilistic causality. On the other hand, the rise of mixed method designs, which seek to balance internal and external validity through a mix of qualitative and quantitative strategies, has partially countered this concern.<sup>3</sup>

The last of these problems, data overload, is a challenge primarily because our time and energy are in limited supply. Yet, with nearly unlimited computing power at our fingertips, this problem has become increasingly manageable. Statistical studies have clearly benefited from advances in computing power, but social scientists have been slower to leverage computing power to improve qualitative or case study research designs.<sup>4</sup>

What we present in this article is a tool that can help both speed, systematize, and assess the process of case selection for scholars seeking to use the comparative method. In this short essay, we introduce a web-based application that can easily identify most and least similar cases from a researcher's dataset. To highlight the need for this application, we begin by reviewing the challenge of case selection in comparative case study research designs. We conclude this essay by illustrating the use of this web-based application to identify cases for comparative analysis.

*'The strength of both statistical and qualitative research designs hinges in large part on the process of case selection.'*

## WHAT TO COMPARE?

The strength of both statistical and qualitative research designs hinges in large part on the process of case selection. A random selection of cases (or a systematic selection of cases that approximates the population) can produce reliable inferential statistics that allow for findings of a sample to be generalized to the larger population. For this reason, large-N statistical studies are better suited to establish the generalizability of findings to a population than are small-N qualitative studies (King *et al*, 1994: 67; Lijphart, 1971: 691), but the reliability of inferences made in statistical analysis is closely linked to the process of case selection.

The strength of small-N research designs is also tied to the processes of case selection. Of course, case selection in small-N research designs is not aimed at generalization to a larger population (George and Bennett, 2004: 30–31; Yin, 2003: 10) through the generation of inferential statistics. Rather, the small-N designs help advance theory by exploring cases that offer a useful combination of representativeness and causal leverage. The identification of cases that offer useful social scientific insights often requires careful reflection on the part of the researcher to pair cases with an effective design or a design with appropriately positioned cases (Gerring, 2007: 144–150).

In particular, comparative method designs (i.e. most similar and most different designs) hinge on the selection of cases that provide the needed variation

across cases on independent, dependent, and control variables.<sup>5</sup> A researcher may be able to readily identify cases that possess the needed variation on the independent and dependent variables, but social scientists are rarely so fortunate to have the desired variation on all relevant control variables as well. The world is not arranged in a way that makes life easy for the social scientist, and cases are rarely available that have the patterns of similarity and difference that would allow for interesting comparisons. Indeed, this is the source of much of the pessimism regarding the comparative method (Glynn and Ichino, 2016; Durkheim, 1982; Mill, 1872). Yet, acknowledging that comparative research designs are limited by the cross-case variation only increases the importance of careful case selection and transparency. A comparative case study design may be imperfect, but there is still much to be gained by selecting cases that produce the strongest design possible.

Scholars employing large-N research designs are able to demonstrate the strength of their designs by clearly laying out the process of case selection. These designs are judged on the extent to which the selection process excludes systematic or research induced sampling bias. For quantitative studies the selection process would, ideally, produce a sample that is a random subset of the population. In small-N studies, researcher bias is more difficult to exclude. The researcher must be intimately involved in selecting cases, giving careful consideration to variation on independent, dependent and control variables.

Compounding matters further, there is no standard guide for identifying, which control variables should be included or prioritized in a selection of cases. What to control for is necessarily a theoretical question. In a symposium on the comparative method Przeworski notes that there is often important confounders that

*'... The world is not arranged in a way that makes life easy for the social scientist, and cases are rarely available that have the patterns of similarity and difference that would allow for interesting comparisons.'*

scholars need to be conscious of when engaging in case selection. He recommends a counter-factual approach to identify potentially complicating dimensions of a comparison, but a counter-factual approach forces the researcher to rely upon existing theoretical understandings to guide her assessment of what is a salient confounding factor and what is of less import (Kohli *et al*, 1995: 18–19).

Nor are there standard guidelines for how differences between cases should be measured or how much weight should be given to each variable (Ragin *et al*, 1996). Other constraints such as data availability, language barriers, and resource limitations further complicate this process of case selection. Thus, the researcher necessarily becomes central to the selection of specific cases for comparison. This, in turn, leaves small-N studies open to the charge of 'cherry picking.'

Fearon and Laitin (2008: 758) describe the problem bluntly but perceptively: 'If one is selecting a few cases from a larger set, why this one and not another? Why shouldn't the reader be suspicious about selection of 'good cases' if no explanation is given for the choice?' This critique can be easily addressed when the total number of cases is quite small. A researcher can describe the criteria for exclusion or inclusion of each potential case. However, when the number of available cases is

large, it is a harder to justify the focus on one pair of cases rather than another.

Concerns over cherry picking can undercut even the most meticulous scholarship. Consequently, case selection is a hugely vexing problem in comparative case study research, and there is no clear answer for how to resolve this problem. Several proposals have been put forward, but there is no consensus on how to proceed. Our task in this essay is not to adjudicate between one approach and another. There are many ways in which case studies can be used, and each research question poses different challenges and opportunities. Rather, we hope to provide scholars who wish to employ comparative case studies as a central or supporting part of their research design with a simple and systematic way of answering Fearon and Laitin's question, 'Why this one and not another?'

### **RECENT PROPOSALS FOR CASE SELECTION**

To avoid the conscious or unconscious 'cherry picking' of cases, there have been multiple attempts to offer more systematic strategies for case selection. Sambanis (2004b) and Gerring (2001) offer innovative strategies for working with within a regression context. These strategies have been used to strong effect by Dafoe and Kelsey (2014) and by DeRouen *et al* (2010). Fearon and Laitin (2008) propose a stratified random selection of cases. However, there has been considerable push back against the idea of random selection for case studies (Freedman, 2008, 4–6; Seawright and Gerring, 2008; Yin, 2003 48; King *et al*, 1994, 124–128), but it is not fully clear how else to proceed. For example, Yin recommends a two-stage process in which the researcher first identifies the pool of relevant cases and then whittles down the

pool by 'defining some relevant criteria for either stratifying or reducing the number of candidates.' This advice encourages a systematic process but does not offer guidance on what a systematic process might look like or how to compellingly communicate that processes to others. In fairness, any attempt to devise an ideal selection processes for case study designs is likely to break down. There is a gap between the ideal and practice in the conduct of research. Contentious researchers are often forced to rely upon systematic but imperfect methods and approaches.<sup>6</sup>

Probably the most comprehensive efforts to develop systematic processes for case selection in situations where a large number of potential cases is available can be seen in the work of Gerring (2007) and Seawright and Gerring (2008). For comparative case study designs, Seawright and Gerring recommend the use of statistical techniques such as propensity matching to identify cases that are assigned similar predicted values by regression models.<sup>7</sup> Yet, this approach implies that similarity rests on a probabilistic logic rather than the necessary and sufficient conditions logic more often associated with the comparative method (Ragin, 2014). This is important because cases might arrive at similar propensity scores through very different mechanisms. Consider a wealthy state with high literacy rates but weak traditions of rule of law and deep ethnic cleavages. This state might have a similar propensity for democracy as a poor state with low literacy rates but ethnic homogeneity and a politics long dominated by the rule of law. From a statistical worldview, these two states might appear quite similar in regard to the likelihood of democratization. Yet, from a necessary and sufficient conditions logic these cases could not be more different. Indeed, the example might be better suited to a most different case design than a most similar

case design. The necessary and sufficient conditions approach to comparative designs expects that cases will align appropriately on each specific dimension. Nielsen (2016) is similarly critical of the propensity scores approach, pointing out that many suggestions for matching encourage case study researchers to adopt a 'statistical world view.'

Nielsen further echoes our concerns that many proposals for matching were developed by statisticians to facilitate large-N analysis and were not designed to extract a small number of cases that would be best suited for further exploration in comparative case designs. The one prominent exception to this approach is Coarsened Exact Matching (see Iacus *et al*, 2012). This approach ensures that cases align on all dimensions by restructuring variables into a limited number of categories. While this approach preserves the logic of the comparative method, it introduces a degree of measurement error. Thus, we believe that there is space for additional strategies and tools that offer both transparent and systematic case selection as well as ease of use for qualitative researchers looking to apply Mill's logics of difference and concurrence.

## **A NEW TOOL AND NEW METRICS**

Gerring (2004) offers a useful typology of case study designs. He notes that case study designs provide variation either over time or across cases or both. The comparative method primarily relies on across case variation (cross sectional or sub-unit). Thus, the effective pairing of cases is essential to an effective comparative method design. Yet, a scholar beginning with a dataset including 194 countries cannot hope to systematically evaluate which cases are actually similar and which are not. The task is simply too gargantuan. Out of 194 countries, we find 18,721

*'... Even when datasets are available, identifying the most similar dyads across three or four variables pushes the limits of what an individual can manage with traditional data management systems.'*

unique dyads (pairs of cases). Even the commonly recommended strategies of selecting cases from within regions (Peters, 1998: 74–79) or within sub-units of a state (Snyder, 2001) do not really simplify the task of comparison. A systematic evaluation of the 50 states in the U.S. would be a significant undertaking. Even when datasets are available, identifying the most similar dyads across three or four variables pushes the limits of what an individual can manage with traditional data management systems. Statistical techniques such as propensity matching or cluster analysis may help, but as noted earlier these techniques were not ontologically aligned with comparative case study research designs and may have steep learning curves for non-quantitatively oriented scholars (see Freedman, 2008: 4).

To help facilitate the process of identifying similar and different dyads from within a dataset, we have developed a web application: the Case Selector (available through the author's faculty webpages at their current institutions, <http://und.edu/faculty/brian.urlacher>).<sup>8</sup> This application compares dyads across a number of user-determined variables for every dyad in a dataset. The application then produces a new dataset with identifying information for the cases in each dyad and measures of similarity for the dependent variable, independent variables, and for control variables. When running the Case Selector, users can

upload their own datasets and input three types of variables. Users may include multiple controls along with independent and or dependent variables. The inclusion of variables is often a function of data availability, but it is also closely related to the purpose of a case study design. As Gerring and Cojocar (2016) argue, a design aimed a hypothesis generation should include a relevant battery of control variables and a dependent variable. A design aimed at assessing a hypothesis would ideally include a solid battery of controls as well as the relevant dependent and independent variables.

### MEASURING DIFFERENCES

The Case Selector follows the practices of earlier computational case selection programs (Nielsen, 2016; Yang *et al*, 2003). Differences between cases are measured using Mahalanobis distances. Mahalanobis distances solve both the question of measuring distance and the weighting of cases in that distances between values for individual cases in the matrix  $X$  for dimensions  $i$  and  $j$ . These linear distances are weighted by the Moore–Penrose pseudo-inverse of the covariance matrix  $S$ , which is composed of the variables being compared.<sup>9</sup> This adjusts the distances (or differences between the values of different cases across variables) to account for variables that may be correlated. See Eq. 1.

$$M(X_i, X_j) = \sqrt{(X_i - X_j)^T S^{-1} (X_i - X_j)} \quad (1)$$

Up to three sets of Mahalanobis distances can be calculated as part of the analysis of a single dataset. When data are available, distances are calculated for the dependent variable ( $d_D$ ), the independent variable(s) ( $d_I$ ), and for control variables ( $d_C$ ). These distances are recorded for each dyadic combination in a dataset and can be used on their own to evaluate cases for appropriateness in a

most similar or most different case study design. Yet, these distances can also be combined into composite scores that allow for a ranking of cases in terms of their appropriateness for most similar and most different designs.

For most similar designs, researchers seek to maximize distances on both the dependent and independent variables with minimal distances on the control variables. Equation 2 translates the three distances into a similarity score. This score is higher when a dyad has properties desirable in a most similar comparative case study design and lower values when there is less divergence in independent and dependent variables or greater divergence in control variables.

$$\frac{\sqrt{d_D} \times \sqrt{d_I}}{d_C} \quad (2)$$

For most different designs, a desirable combination of cases will have nearly identical values for both the dependent and independent variables but will be highly divergent on all control variables. A difference score is provided in Eq. 3. This difference score takes on greater values when the numerator, which is simply the Mahalanobis distance for control variables, is large and the denominator is small, which occurs when the Mahalanobis distances for both the independent and dependent variables are small.

$$\frac{d_C}{d_D + d_I} \quad (3)$$

### POTENTIAL APPLICATIONS

The Case Selector produces a list of all the possible dyads in a dataset, along with the differences between cases on a dependent, independent, and control variables. Gerring and Cojocar (2016) note that depending on the purpose of a comparative case study design a

researcher might wish to focus on different pairings of control, independent, or dependent variables. The Case Selector facilitates this process with graphing options. In addition to being able to easily graph the differences between pairs of cases, the Case Selector allows for all comparisons to be graphed or a subset of dyads involving a specific case. So, how might this data be used? There are at least three ways in which this information can aid in the process of case selection: (1) identifying suitable dyads for further study, (2) identifying a good match for a case already selected, and (3) evaluating a pair of pre-selected cases.

To identify suitable dyads for further study, a researcher would begin by identifying as broad a sample of potential cases as possible. Armed with this list, the researcher should identify potentially relevant control variables. After assembling a dataset of relevant cases and relevant variables (or identifying a pre-existing dataset), the researcher can load this dataset into the Case Selector.<sup>10</sup> Variables for inclusions should be entered and output generated.<sup>11</sup> After running the Case Selector and generating comparisons of all possible dyads, the researcher can select the most similar (or most different) cases for preliminary investigation. This preliminary investigation might be aimed at identifying cases that have the needed variation on variables that are not included in the dataset, or it might be aimed at judging the feasibility of studying specific cases.

To identify a good match for a case already selected, the same process would be followed. When sorting the output from the Case Selector, the researcher would first separate out dyads that contain the case already selected for study. Within this sub group the researcher can further whittle the list down by sorting for most similar (or most different) cases.

To evaluate a pair of cases already selected, the researcher would generate dyad comparisons, sort the dyads, and

then identify the rank of the pre-selected dyad within the larger dataset of dyads. This allows for the researcher to evaluate his or her case selection. If a dyad ranks in the top 10 percent of most similar dyads (or most different, depending on the desired design), then this would make for a stronger design than a selection of cases in the top 20 percent of dyads. Being able to precisely communicate where a combination of cases falls within the universe of possible combinations is a critical piece of information for addressing the cherry-picking charge.

## AN ILLUSTRATION

Each of the three uses of the Case Selector that we have proposed above is illustrated here using two recent studies of civil war violence against civilians. Reed M. Wood's (2010) statistical analysis of violence against civilians provides a pool of cases. Jeremy Weinstein's (2007) book, *Inside Rebellion*,<sup>12</sup> provides a comparative case design that we evaluate and supplement with the Case Selector.

Weinstein's study provides an excellent application of the most similar case comparative method. For his analysis, he selects two pairs of rebel groups for study. The first pairing is between Uganda's National Resistance Army (NRA) and Mozambique's Renamo. The second pairing involves two factions of Sendero Luminoso in Peru. Weinstein notes divergent behaviour, particularly in the use of violence against civilians, and identifies a potential cause: the availability of economic and social resources that shapes the organizational development of rebel groups.

As with all matching procedures, the first step is to work through key conceptual and theoretical aspects of the research question. Basic questions related to the potential scope of a phenomena are critically important to address. In practice,



scope questions often get resolved by the structure of existing datasets. In the analysis of Weinstein's case selection, we draw on the Uppsala Conflict Data Program's (UCDP) data. This decision imposes temporal and conceptual limitations, but these limitations are reasonably well understood and the potential consequences of case selection in civil wars has been debated and discussed (Sambanis, 2004a; Kuperman, 2004).

A second conceptual challenge that researchers must resolve is the selection of variables on which to match. As with model specification in a quantitative context, theoretically salient variables should not be excluded from a model, but there are also undesirable consequences of deploying 'kitchen sink models' (Schrodt, 2014). This problem is no less salient in a qualitative context where researchers must identify what dimensions of similarity or difference are salient. The inclusion of theoretically irrelevant variables can have the effect of eliminating otherwise viable comparisons. However, the failure to incorporate theoretically salient variables can yield comparisons that make for a weak comparative design.

For this reason, one might conclude that the specification of selection criteria is potentially even more critical in comparative case study designs than statistical models. Without an error term, algorithm based selections processes have no easy way to incorporate the uncertainty of stochastic processes or the effect of factors not explicitly incorporated into an analysis. It is here that the comparative method research must necessarily turn back into the realm of the researcher's judgement. Identified a set of criteria for selection does not serve as a substitute for rigorous knowledge of a topic and at least rudimentary knowledge of the details of specific cases. Researcher's still need to identify and weigh potentially relevant factors that were either not incorporated into the initial selection

process or were poorly measured. Thus, we stress that the Case Selector (or any other selection algorithm) should be viewed as a tool for managing complexity and not as the case study equivalent of regression output.

Rather than attempting to argue for the inclusion or exclusion of specific variables, we defer to the literature on civilian targeting in civil war.<sup>13</sup> We draw on a statistical model of rebel group violence against civilians developed by Reed M. Wood (2010) to guide our decision on which variables to include.<sup>14</sup> Wood uses the UCDP one-sided violence data and tests a wide range of competing hypotheses that have been offered to explain the use of violence against civilians. Wood (2010) provides a detailed discussion of the operationalization of these variables in his article, so we will not discuss operationalization here.

A third conceptual problem to address is how to handle the temporal aspect of panel data. Data that is organized around a country-year or a conflict-year, often has multiple observations per case. In a regression context, these additional cases provide useful information. Similarly, in case study designs temporal information often supplies useful variation; however, the primary source of variation in comparative designs derives from the across case comparisons. Identifying across case variation can be obscured when there are multiple observations of the same case. It is rarely useful from the perspective of the comparative method that the closest match to a case is that same case in the preceding year.

Thus, some technique is needed to collapse the temporal information in datasets. An averaging of variables across time might make sense in situations where change over time is not theoretically salient.<sup>15</sup> Alternatively, researchers might opt for data from the year before the start of temporal processes. If researchers do seek to



incorporate time as a salient feature of a comparative design, we recommend considering two approaches for highlighting temporal variation in a way that reduces the multiple observations per case. First, researchers might incorporate the observed difference within each case between the minimum and maximum values on variables of interest. Second, a researcher might take the difference between a start point of some processes and an end point, essentially a pre-post treatment comparison. We offer no prescriptions for how researchers should approach the problem of case selection from panel data beyond the advice that researchers should be guided by theory and that researchers should be transparent in the decisions that they make.

To this end, we opted to collapse temporal information in Wood's data by averaging. Many of the variables are either static over time (conflict area, density, conflict type, and availability of lootable resources). Others are relatively slow to change (conflict duration, log of GDP per capita). Yet, some variables, particularly those related to conflict severity and the use of violence by the government, have potentially important variation over time. Before adopting a specific pairing of cases, a researcher should examine the temporal patterns for potentially salient shifts in these variables over time.

### **IDENTIFYING MOST SIMILAR CASES**

This analysis of the UCDP one-sided violence data includes 179 different rebel groups, which produces 15,931 dyads for comparison. To identify cases that would be strong candidates for a most similar case with a different outcomes research design, we began by sorting dyads from smallest to largest in terms of the similarity score. Table 1 presents strong candidates for a most similar case design.

Several of the dyads in Table 1 involve two rebel groups from the same country. A number of other dyads are geographically proximate. This partially validates two strategies that comparative researchers have long used to control for differences between cases, namely comparing units within a single country (Gerring, 2004: 348; Snyder, 2001) and looking within regions (Dogan, 2009: 23; Lijphart, 1971: 688) for similar cases.

### **IDENTIFYING SUITABLE MATCHES**

Weinstein observes two rebel factions within a single conflict in Peru. As shown above, this can be a powerful design, but it hinges on there being multiple groups or sub-units that can be observed that also have divergent outcomes. Had a split in Sendero Luminoso not occurred, Weinstein would potentially have needed to identify an additional case for comparison. When a case has already been selected, the Case Selector can aid in identifying a useful case for comparison. The same output file generated for the previous example, when manipulated in a slightly different way, can provide guidance in this process. We began looking for cases for comparison to Sendero Luminoso by selecting only dyads that include Sendero Luminoso. Within these dyads, we sorted dyads according to the similarity score from largest to smallest.

Again, geographic proximity seems to work as a potential control strategy for a wide variety of factors. Sendero Luminoso is quite similar to a number of other Latin American rebel groups including FARC, EPL, and ELN in Columbia, URNG in Guatemala, and FMLN in El Salvador. Unfortunately, for the purposes of case selection, these cases are also quite similar in terms of observed violence against civilians. To find comparable cases that offers the needed variation on the

**Table 1: Most similar cases with most different outcomes**

Case 1	Case 2	Mahalanobis distances		Similarity score
		Dep. var.	Ctrl. vars.	
D.R. Congo: MLC	D.R. Congo: RCD	0.516	0.028	18.733
Bosnia: Croatian Irregulars	Bosnia: Serbian Irregulars	7.38	0.472	15.646
Moldova: Dniestr Republic	Bosnia: Serbian Irregulars	7.543	1.546	4.877
Georgia: Republic of Abkhazia	Bosnia: Serbian Irregulars	7.451	1.576	4.729
Bosnia: Serbian Irregulars	Azerbaijan: Nagorno-Karabakh	7.543	1.658	4.549
Bosnia: Serbian Irregulars	Sudan: SAF	7.543	1.828	4.126
Philippines: MNLF faction	Bosnia: Serbian Irregulars	7.543	1.955	3.857
D.R. Congo: AFDL	Sudan: JEM	8.847	2.881	3.071
Bosnia: Serbian Irregulars	Yugoslavia: Croat irregulars	7.543	2.473	3.05
Philippines: ASG	Bosnia: Serbian Irregulars	7.481	2.46	3.041

dependent variable, a geographically broader net needs to be cast. Figure 1 provides an illustration of the proximity of select rebel groups to Sendero Luminoso, which is located at the origin in Figure 1. Three cases (Renamo, JVP, and the Khmer Rouge) cluster close to Sendero Luminoso near the origin.

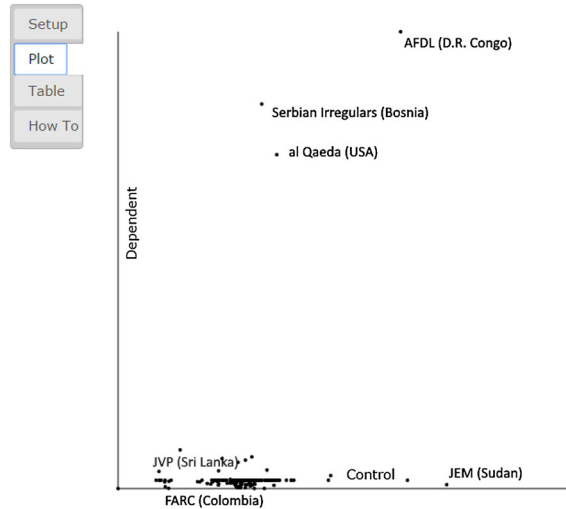
Along the horizontal axis, there are three cases that would be potential candidates for a most different case research design. The National Salvation Front, JEM, and UIFSA are highly divergent cases from Sendero Luminoso in terms of the 10 control variables identified in Wood's analysis. These cases are also quite similar to Sendero Luminoso in terms of one-sided violence.

The upper left corner of Figure 1 is where ideal pairings would be located for a most similar case design. There are three rebel groups that are potential contenders (AFDL, Serbian Irregulars, and UDCA/LRA). While the UDCA/LRA case is the most similar to Sendero Luminoso, it does not have as large a divergence on the dependent variable as is seen in both the AFDL and Serbian Irregular cases. These two cases,

however, are a less good match in terms of control variables. While a researcher might opt to investigate one or all of the three cases in the upper left corner of the graph, the Case Selector aids in promoting transparency in the selection process by giving researchers a way to demonstrate the trade-offs inherent in selecting a workable comparative case.

### **EVALUATING PREVIOUSLY SELECTED CASES**

A final potential application of the Case Selector is to evaluate cases that have already been selected. Data availability (or non-availability), access to informants, financial limitations, language skills, or security concerns may restrict the options researchers have for selecting cases. While this may not be ideal, it should not be assumed that comparative case study designs selected for practical reasons will automatically be weaker than cases selected more systematically. The degree of similarity or difference between cases is an empirical question and should be resolved with data.



**Figure 1** Comparison of select cases with Sendero Luminoso.

To illustrate how this might work, we show how a single dyad compares against the entire range of potential dyads. In particular, Weinstein’s comparison of Renamo in Mozambique and the NRA in Uganda is examined. Within the larger pool of 15,931 dyads, the NRA-Renamo dyad ranks 1708 for similarity in terms of control variables. This translates to 10.7 percent of dyads being more similar than the NRA-Renamo dyad and 89.3 percent of dyads being less similar. The NRA-Renamo dyad ranks 14,998 for the dependent variable when sorting from most similar to least similar. Thus 5.98 percent of dyads are more different in their outcomes than the NRA-Renamo dyad.

The NRA-Renamo dyad stacks up quite well against the pool of dyads available for study. The dyad achieves a relatively high level of similarity in terms of control variables and a notably high level of difference for the dependent variable. While there might be dyads that would offer greater control with similar levels of divergence on the dependent variable, the selection of the NRA-Renamo dyad would certainly be a defensible selection given the pool of dyads available.

## CONCLUSION

This article has wrestled with a persistent problem in comparative case study design: identifying which cases to compare. This has been a long-standing challenge in the conduct of comparative case study research, but it is particularly relevant given the emerging consensus around value multi-method designs (Mahoney, 2010: 138).<sup>16</sup> The Case Selector is one tool in a growing toolbox available to case study researchers to manage the information overload that occurs when selecting a strong combination of cases large number of potential cases. By no means do we believe that the Case Selector will be the definitive or even optimal solution to the problem of case selection. The Case Selector adds to the toolbox of available techniques, which includes propensity scores, coarsened exact matching, and others. Each proposed method necessarily contains limitations and challenges. Still, a more flexible and user-friendly toolbox of case selection techniques is critical in promoting greater transparency in the process of case selection.

Utilizing the Case Selector, or a similar technique, encourages researchers to declare explicitly which controls are used and how they are measured. While this is a very basic element of case selection, it is often glossed over in the communication of case study design.<sup>17</sup> Researchers can also describe more precisely how a dyad compares against other possible dyads. Precise statements about what percentage of cases are more or less similar in terms of independent, dependent, and control variables can help to assuage concerns that cases were 'cherry picked' by the researcher and thus should be treated as suspect. We see clear metrics of similarity and difference as vital to communicating strong comparative case designs, and we see this as our primary contribution to the comparative method. The development of similarity and difference scores as described in this project could greatly enhance the assessment of case selection, particularly when data exists for all relevant control, independent, and dependent variables.

The Case Selector has several advantages related to accessibility over other approaches to case selection and identification that have been proposed. First, the application was designed for case study researchers. This is not an added function to an existing statistical package, nor is it a statistical technique that can be re-worked to provide information useful in case selection. Rather, this web application is designed specifically to provide information to researchers looking for most similar and most different dyads. Consequently, the learning curve for the Case Selector is significantly reduced. Second, the method is designed to be user friendly. The Case Selector is built as a webpage that allows intuitive 'drag and drop' placement of variables and drop down menus for selecting options.

Of course, the Case Selector would not be an appropriate tool to use in every situation. If the pool of potential cases is

*'... Precise statements about what percentage of cases are more or less similar in terms of independent, dependent, and control variables can help to assuage concerns that cases were 'cherry picked' by the researcher.'*

quite small, the researcher could perform a systematic comparison of dyads without the use of the Case Selector. Alternatively, a researcher might opt to study all of the cases available. In addition, a careful application of the Case Selector requires that a well-developed dataset be available or that it can be created. The Case Selector may not be useful in the initial phase of a research program when data collection efforts are in the early stages. On the other hand, research programs that have been underway for years are likely to have well developed datasets available that cover many theoretically salient control variables. Under these conditions, the Case Selector can help facilitate more careful case selection and ideally can help to improve the usefulness of the comparative case study research design as a tool in the social scientist's methodological toolbox. Indeed, the Case Selector has potential to open up new lanes of research on problems that have been deeply explored statistically, but for which systematic case study work has trailed behind.

## **Acknowledgements**

The authors would like to thank the anonymous reviewers for their insightful comments and feedback over the course

of the review processes. This project has been significantly improved by their suggestions. The authors have also agreed to provide access to the Case Selector through their faculty webpages at their affiliated institutions.

## Notes

1 Glynn and Ichino (2016) argue that there are subtle but important distinctions between Lijphart's (1971) framing of the comparative method and the logic of similarity and difference outlined by Mill (1872).

2 Much of this debate over qualitative methods can be seen in the reaction of scholars to King *et al* (1994) *Designing Social Inquiry*, particularly in Brady and Collier's (2004) edited volume *Rethinking Social Inquiry*.

3 Slater and Ziblatt (2013) describe a trend within the field that attributes external validity to large-N approaches and internal validity to comparative or small-N methods. While they argue that this division is certainly possible, they also document the sophisticated use of small-N methods to generalize trends documented within single cases through statistical techniques. They ultimately conclude that the comparative method remains a robust and versatile tool in the social scientist's toolbox.

4 This is not to say that progress has not been made. Work by Nielsen (2016) has paved the way forward in the development of algorithmic approaches for case selection. While Gerring and Cojocar (2016) are of mixed mind about the value of algorithmic case selection, they recognize movement in this direction.

5 See Przeworski and Teune (1970) for a clear discussion of the logic underlying these two variations of the comparative method.

6 It should be noted, that this is not unique to qualitative case selection. While sampling is well understood, most public opinion research does not rely upon Simple Random Samples (SRS). The Random Digit Dialling (RDD) approach is a flawed approximation of a SRS. The limits of this method are known and different polling firms take steps to correct for limitations of RDD (Asher 2016). Yet, there is no optimal solution. Thus, case selection for both case study designs and public opinion polling is best understood as a mix of systematic method and art.

7 A similar strategy is proposed by Sambanis (2004b), who calls for case study research designs that focus on cases predicted well and predicted poorly by regression models.

8 The Case Selector is primarily a tool for comparative (most similar and most different) designs. The data generated through this tool is not structured to facilitate other types of case study designs. To select crucial cases, extreme cases, or typical cases the techniques outlined by Gerring (2001) may be more useful.

9 Mahalanobis distances traditionally use the *inverse* of the covariance matrix. This does not exist if the data is linearly dependent. The pseudo-inverse does exist, and is equivalent to the inverse along the subspace of independent data.

10 The Case Selector uses the Comma Separated Variable (.csv) format. This is a standard format that most statistical packages can accommodate for either import or export.

11 For detailed instructions on how to manipulate the options available with the Case Selector, see the Case Selector codebook and tutorials, which are available with the application.

12 It should be stressed that we are not seeking to second guess or critique the appropriateness of Weinstein's case selection. Weinstein provides a solid justification for his case selection in his book, and indeed his diligence is largely supported in this illustration. Still, there is value in revisiting Weinstein's case selection (indeed only good can come from scrutinizing and assessing case selection). We also are able to offer suggestions for other pairings of cases that might compliment Weinstein's case selection either in terms of the most similar or most different method.

13 In recent years, several quantitative studies have sought to explain the use of violence against civilians within a single conflict (Balcells, 2010; Kalyvas, 2006) or across multiple conflicts (Wood, 2010; Eck and Hultman, 2007).

14 One change to the data used by Wood (2010) is the inclusion of an additional case: Uganda's National Resistance Army. Although this case was not included in UCDP's data, the rebel group was one of the cases studied by Weinstein and needs to be included for comparative purposes. Wood's coding procedures were followed in coding the additional case. Data on rebel and government violence against civilians in Uganda is provided in Weinstein's (2007) book.

15 In a statistical analysis, this kind of aggregation would be highly problematic, as one of the central elements of causality is that the cause precedes the effect. For matching purposes, however, case selection is often an iterative process. Averaging might be useful as a first stage in a larger processes of selection. A researcher could follow this initial selection with a more focused analysis that considers any large shifts in variables over time that might be problematic.

16 While Mahoney notes the rise of multimethod designs that combine qualitative and quantitative methods, this trend is not universally embraced. Ahmed and Sil (2012) for example argue that single method research designs better allow for methodological pluralism in part because they avoid the epistemological closure that inadvertently results from the methodological hegemony of quantitative approaches.

17 See Maoz (2002: 164) for a biting articulation of this critique.

## References

- Ahmed, A. and Sil, R. (2012) 'When Multi-method Research Subverts Methodological Pluralism—or, Why We Still Need Single-Method Research', *Perspectives on Politics* 10(4): 935–953.
- Asher, H. (2016) *Polling and the Public: What every citizen should know*, Washington, DC: CQ Press.
- Balcells, L. (2010) 'Rivalry and Revenge: Violence Against Civilians in Conventional Civil Wars', *International Studies Quarterly* 54(2): 291–313.
- Brady, H. E. and Collier, D. (eds) (2004) *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Lanhan, MD: Rowman and Littlefield.
- Dafoe, A. and Kelsey, N. (2014) 'Observing the capitalist peace: Examining market-mediated signaling and other mechanisms', *Journal of Peace Research* 51(5): 619–633.
- DeRouen Jr, K., Ferguson, M. J., Norton, S., Park, Y. H., Lea, J., and Streat-Bartlett, A. (2010) 'Civil War Peace Agreement Implementation and State Capacity', *Journal of Peace Research* 47(3): 333–346.
- Dogan, M. (2009) 'Strategies in Comparative Sociology', *New Frontiers in Comparative Sociology*, Leiden, Netherlands: Brill, pp. 13–44.
- Durkheim, E. (1982) [1895] *The Rules of Sociological Method*, New York, NY: Free Press.
- Eck, K and Hultman, L. (2007) 'One-Sided Violence Against Civilians in War: Insights from New Fatality Data', *Journal of Peace Research* 44(2): 233–246.
- Fearon, J. D. and Laitin, D. D. (2008) 'Integrating Qualitative and Quantitative Methods', *The Oxford Handbook of Political Methodology*, New York, NY: Oxford University Press. pp. 756–776.
- Freedman, D. A. (2008) 'Do the Ns Justify the Means?', *Qualitative & Multi-Method Research* 6(2): 4–6.
- George, A. L. and Bennett, A. (2004) *Case Studies and Theory Development in the Social Sciences*, Cambridge, MA: The MIT Press.
- Gerring, J. and Cojocaru, L. (2016) 'Selecting Cases for Intensive Analysis: A Diversity of Goals and Methods', *Sociological Methods & Research* 45(3): 392–423.
- Gerring, J. (2001) *Social Science Methodology: A Criterial Framework*, Cambridge, UK: University of Cambridge Press.
- Gerring, J. (2004) 'What Is a Case Study and What Is It Good for?', *American Political Science Review* 98(2): 341–354.
- Gerring, J. (2007) *Case Study Research: Principles and Practices*, Cambridge, UK: Cambridge University Press.
- Glynn, A. N., and Ichino, N. (2016) 'Increasing Inferential Leverage in the Comparative Method Placebo Tests in Small-n Research', *Sociological Methods & Research* 45(3): 598–629.
- Iacus, S. M., King, G., Porro, G., and Katz, J. N. (2012) 'Causal Inference Without Balance Checking: Coarsened Exact Matching', *Political Analysis* 20(1): 1–24.
- Kohli, A., Evans, P., Katzenstein, P. J., Przeworski, A., Rudolph, S. H., Scott, J. C., and Skocpol, T. (1995) 'The Role of Theory in Comparative Politics: A Symposium', *World Politics* 48(1), 1–49.
- Kalyvas, S. N. 2006. *The Logic of Violence in Civil War*, Cambridge, UK: Cambridge University Press.
- King, G., Keohane, R. O., and Verba, S. (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton, NJ: Princeton University Press.
- Kuperman, A. J. (2004) 'Is Partition Really the Only Hope? Reconciling Contradictory Findings About Ethnic Civil Wars', *Security Studies* 13(4): 314–349.

- Lijphart, A. (1971) 'Comparative Politics and Comparative Method', *American Political Science Review* 65(3): 682–698.
- Maoz, Z. (2002) 'Case Study Methodology in International Studies: From Storytelling to Hypothesis Testing', *Evaluating Methodology in International Studies*, Ann Arbor, MI: University of Michigan Press, pp. 161–186.
- Mahoney, J. (2010) 'After KKV: The New Methodology of Qualitative Research', *World Politics*, 62(1): 120–147.
- Mill, J. S. (1872) *A System of Logic*. London, England: Longmans, Green, Reader, and Dyer.
- Nielsen, Richard A. 2016. Case Selection via Matching. *Sociological Methods & Research* 45(3): 569 - 597.
- Peters, B. G. (1998) *Comparative politics: Theory and methods*, Washington Square, NY: New York University Press.
- Przeworski, A. and Teune, H. (1970) *The Logic of Comparative Social Inquiry*, New York, NY: Wiley-Interscience.
- Ragin, C. C. (2014) *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.
- Ragin, C. C., Berg-Schlosser, D., and de Meur, G. (1996) 'Political Methodology: Qualitative Methods', *A New Handbook of Political Science*, New York, NY: Oxford University Press, pp. 749–769.
- Sambanis, N. (2004a) 'What is Civil War? Conceptual and Empirical Complexities of an Operational Definition', *Journal of Conflict Resolution* 48(6): 814–858.
- Sambanis, N. (2004b) 'Using Case Studies to Expand Economic Models of Civil War', *Perspectives on Politics* 2(2): 259–279.
- Seawright, J. and Gerring, J. (2008) 'Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options', *Political Research Quarterly* 61(2): 294–308.
- Schrodt, P. A. (2014) 'Seven Deadly Sins of Contemporary Quantitative Political Analysis', *Journal of Peace Research* 51(2): 287–300.
- Slater, D. and Ziblatt, D. (2013) 'The Enduring Indispensability of the Controlled Comparison', *Comparative Political Studies* 46(10): 1301–1327.
- Snyder, R. (2001) 'Scaling Down: The Subnational Comparative Method', *Studies in Comparative International Development* 36(1): 93–110.
- Tarrow, S. (2010) 'The Strategy of Paired Comparison: Toward a Theory of Practice', *Comparative Political Studies* 43(2): 230–259.
- Weinstein, J. M. (2007) *Inside Rebellion: The Politics of Insurgent Violence*, Cambridge, UK: Cambridge University Press.
- Wood, R. M. (2010) 'Rebel Capability and Strategic Violence Against Civilians', *Journal of Peace Research* 47(5): 601–614.
- Yang, Z., Matsumura, Y. Kuwata, S., Kusuoka, H., and Takeda, H. (2003) 'Similar Case Retrieval from the Database of Laboratory Test Results', *Journal of Medical Systems* 27(3): 271–281.
- Yin, R. K. (2003) *Case Study Research: Design and Method*, Thousand Oaks, CA: Sage Publications, Inc.

## About the Authors

**Timothy Prescott** is an Associate Professor of mathematics at the University of North Dakota. His research focuses on probability, and he also programs in Python, Javascript, and LaTeX in his spare time.

**Brian R. Urlacher** is an Associate Professor of political science at the University of North Dakota. His research focuses on conflict resolution, international cooperation, and negotiation theory. His work in these areas has appeared in *International Studies Quarterly*, *International Negotiation*, and *Civil Wars*. His book *International Relations as Negotiation* was published in 2015 by Routledge Press.