

Research Statement

Gustavo Diaz

Department of Political Science

Northwestern University

gustavo.diaz@northwestern.edu

Social scientists often study phenomena that cannot be observed directly. For example, we use responses to hypothetical survey questions to infer actual behavior, we resort to aggregate election results to understand individual evaluations of politicians' performance in office, and we conduct randomized controlled trials in some places to determine if a policy is advisable in other places. Doing this credibly requires careful research design, since researchers must try to anticipate the challenges to inference before conducting data analysis.

My research develops standards to navigate the tradeoffs that emerge as one considers research design alternatives before data collection. I use tools from design-based causal inference and data science to identify practices and procedures that researchers can adopt to improve how they approach data collection at the pre-analysis stage. My current agenda focuses on improving statistical precision, since this is often the deciding factor when choosing among alternative research designs, all promising unbiased estimators.

I apply the insights of my methodological work to questions in international politics and political behavior. The common theme across these applications is the goal of improving our ability to make statistical inferences about hard-to-observe social and political phenomena, which in turn leads to a more credible evidence base to support decision-making.

Improving precision before data collection

The last decade has seen considerable improvement in research transparency and registration. Recent advances in experimental design provide tools to diagnose the properties of a research design before data collection. For example, one can think about bias, power, or target sample size under different hypothetical data generation processes.

A recurrent goal in statistics, econometrics, and social science methodology is to minimize bias, or being close to the hypothetical truth on average. When planning an original data collection effort, researchers can often choose among many alternative research designs, all with previously identified unbiased estimators. My agenda focuses on optimizing statistical precision, understood as producing consistent results after multiple realizations of the same data generation process. While this is a crucial factor when choosing a research design, the literature implicitly assumes that one can simply improve precision by increasing sample size, without providing much guidance to assess among alternative designs. Moreover, simply increasing sample size is not feasible in most applications due to practical or ethical considerations. Therefore, even when one does not face a choice between alternative designs, statistical precision is still paramount.

Focusing on the design and analysis of survey and field experiments, my agenda advances the argument that the choice between alternative research designs for which unbiased estimators are already documented is rarely free. A design promising improvements in statistical precision without sacrificing unbiasedness often brings unforeseen costs in other dimensions.

Two projects exemplify how I develop standards to navigate research design choices with unforeseen challenges. First, in an article recently published in *Political Analysis* with Erin Rossiter (Notre Dame), we discuss the circumstances under which adopting research design features aimed at improving precision can instead hurt it through implicit or explicit sample loss. For example, block randomization can improve precision, but if this requires contacting participants multiple times to collect pre-treatment blocking covariates, then it creates space for attrition that would not exist otherwise, which may offset the precision gains from blocking. We posit this is the main reason why researchers deviate from the standard experimental design infrequently. Through three replications and six reanalyses of previously published experiments in leading political science journals, we show how precision gains from alternative designs can withstand significant degrees of sample loss. From this exercise, we also identify guidelines to navigate the tradeoff between precision and sample retention in experiments.

Second, in a manuscript published at the *Journal of Experimental Political Science*, I propose statistical tests to address problems with double list experiments. Social scientists use list experiments in surveys when respondents may not answer truthfully to sensitive questions. When their assumptions are met, list experiments reduce sensitivity biases from misreporting. However, they tend to produce estimates with high variance, which prevents researchers from improving upon direct questioning. Double list experiments promise to remedy this by implementing two parallel list experiments and then aggregating their results, which roughly halves the variance of the estimate for the prevalence of the sensitive trait.

This implies an estimator that is more precise and still unbiased, but their implementation brings the question over whether the aggregation of the results of two parallel experiments yields a valid estimate. Using a reanalysis of a study on support for anti-immigration organizations in California as a running example, the tests leverage variation in the order in which respondents see the sensitive item to detect whether respondents are reacting to list experiment questions in unintended ways. This provides researchers with a tool to apply this underexplored variant of the technique more widely.

This agenda is currently extending toward improving precision by combining different techniques that target the same quantity of interest, and it has also let me to contribute to research in regions beyond my expertise. For example, in work embedded on a recently granted an *European Research Council Starting Grant* with Inés Fynn (Universidad Católica del Uruguay), Verónica Pérez (Universidad de la República), and Lucía Tiscornia (University College Dublin), we combine list experiments and the network scale up method (NSUM), a popular technique in the health sciences, to improve precision in the estimation of the prevalence of sensitive attitudes and behaviors. Previous work combining different techniques to minimize sensitivity bias in survey questions relies on asking direct questions, altered research designs, cumbersome

statistical modeling assumptions, or access to population-level data, all of which are problematic in their own way. By using NSUM questions as auxiliary information to the list experiment, we manage to improve precision by only imposing one additional assumption: that people with disproportionately high exposure to the sensitive trait of interest in their personal network are likely to hold the trait themselves. This does not apply to every sensitive attitude or behavior of interest in the social sciences, but is less demanding than the assumptions required to generalize NSUM estimates to a target population.

Substantive contributions

My expertise has also translated into substantive contributions in a broad range of applications. I have primarily focused on issues of accountability, governance, and representation in the Global South. These contributions fall into two categories.

First, I have focused on projects that help illuminate survey experimental research practice. For example, in a manuscript forthcoming at the *British Journal of Political Science* with Virginia Oliveros (Tulane), Rebecca Weitz-Shapiro (Brown), and Matt Winters (Illinois), we use a survey experiment in Argentina to study gendered differential reactions to policy implementation. Previous work suggests that women face higher scrutiny for their performance in office. However, in the context of the implementation of a food distribution program, we find that voters are only responsive to performance information among men officeholders and tend to ignore performance information when told that an officeholder is a woman. We attribute this result to voters' perception of men politicians as the default category, thus providing no new information on top of performance. In turn, mentioning the gender of a woman politician leads voters to believe that good performance stems from factors beyond the incumbent's control. This has broader implications for research on gender gaps in the evaluation of public officials, since most survey-experimental analysis on the topic fail to account for this informational imbalance.

A second set of projects focuses on applying methodological innovations to improve statistical precision in areas where researchers need it the most. For example, in a working paper with the same team that proposes combining list experiments with the NSUM, we document the prevalence of criminal governance strategies in Uruguay. This is a country with criminal organizations that are small in size but still deeply rooted in pockets of the capital city. Here, estimating the extent of exposure to criminal governance is challenging due to the combination of sensitivity bias and our inability to oversample on key subgroups of interest, which means any attempt to maximize statistical precision needs to come from our research design choices. This is the basis of a recently awarded *European Research Council Starting Grant* seeking to understand criminal governance in least-expected contexts from a comparative perspective, which will in turn serve as a platform for further methodological innovation in survey research.