Personal Statement

Gustavo Diaz
Department of Political Science
Northwestern University
gustavo.diaz@northwestern.edu

Research agenda

Political 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 even before conducting data analysis.

My research develops standards to navigate the tradeoffs that emerge when as considers research design alternatives before data collection. I use tools from statistics and computational social 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 around the challenges to accountability and representation in political economy 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. For the sake of brevity, this document focuses on my methodological agenda only.

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 surveys and 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 under unforeseen challenges. First, in a manuscript *conditionally accepted at 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. 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. For example, in work in progress 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 disproportionally 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.

Teaching philosophy and experience

As an instructor of quantitative methods for political scientists, I face a polarized audience. Increasingly, students start their program with considerable experience in mathematical thinking, statistics, and statistical programming. Still, many start with an appreciation for data analysis, but come from educational and career paths designed to explicitly avoid math.

My approach to keep both audiences engaged within one term is to unify math, statistics, and coding as the task of acquiring a new language. A single course will not teach students everything they need to know to become fluent, but it can give enough tools to facilitate future learning in a direction that is beneficial to students regardless of their background and career goals. For some, this may mean engaging directly with data and code or even creating new methodologies. For others, the goal may be just to communicate productively with scholarship drawing on quantitative findings or data analysts at the workplace.

To accommodate this diversity, I design courses with two principles in mind. First, students need flexibility to engage with the course on their own terms and focus on the content they find useful. For example, the flipped classroom lab sessions in my course on data analysis for public opinion and policy at McMaster asked students to evaluate a research design, suggest alternatives or modifications, and to evaluate its statistical properties through coding and writing. Some students may propose increasing the sample size, sampling from a different underlying population, or changing the assignment of treatment conditions. This allows students to pursue the tasks that suit their interests and gives the instructor freedom to reward creativity and effort over correctness.

The flexibility principle also applies to the problem sets in my graduate introductory methods course, where contract grading allows me to reward learning even when students get stuck with coding. I place equal value on reasoning why something did not work and how it should look had the code worked properly as I do on producing working code. This reduces anxiety around finding the right answers, and encourages creativity and collaboration.

The second principle is accountability, which is necessary to keep everyone on task while allowing flexibility. This means agreeing on an overarching theme that every single course activity must relate to. For example, early on my data analysis for public opinion and policy course, I introduce the bias-variance tradeoff as a principle to choose among alternative research designs. So, while students are free to propose any modification to an existing research design that they

deem appropriate, they are also required to identify the explicit or implicit costs associated with their proposal. They must consider, for instance, that a representative sample is more expensive than a convenience sample, or that implementing a block-randomized experiment may require access to variables that cannot be measured easily.

Similarly, the overarching theme in my graduate probability and statistics course is how assumptions shape the inferences that we can credibly draw from data. I emphasize how we need to make unrealistic assumptions, even if minimal, to enable statistical inference, and that we need to hold ourselves accountable to those assumptions when evaluating the appropriateness of a statistical procedure.

Flexibility and accountability also help in preventing instances of discrimination in the learning process. Through flexibility, students are invited to add value to the course by bringing their own perspective, knowledge, or experiences. In turn, accountability sets the scope for the type of contributions of interventions that are admissible. From this perspective, a racist remark is unacceptable not because someone disagrees with it, but because it is beyond the scope of the vocabulary we aim to build.

In terms of experience, I am the central person teaching statistics courses at Northwestern political science. I teach the first course in the graduate methods sequence, focusing on probability and statistical inference. At the undergraduate level, I also teach the introduction to research methods in political science, a course that teaches political science majors to become informed consumers of data analysis and prepares them for more advanced courses, and a research seminar on the use of experiments and machine learning to inform decision-making in academia, policy, and industry. I also lead the math camp program for incoming political science and sociology students, and run the year-long statistical computing workshop that introduces cutting-edge coding practices. Next year, I will add a graduate seminar on machine learning to my portfolio.

Before joining Northwestern, I taught data analysis for public policy and public opinion at McMaster, with emphasis on using quantitative evidence to make credible policy recommendations. The goal of this course is to give students hands-on experience in designing quantitative research projects in an area relevant to academia, policy, or industry.

At Tulane, I taught an undergrad senior course on the challenges that developing democracies face from the perspective of evidence-based policy making. This course overviews the main challenges in the path to democratic consolidation around the world, proposed solutions to these challenges, and how governments, researchers, and civil society organizations use data to evaluate these solutions. The previous version focused primarily design-based causal inference, but future versions will also feature data science and machine learning.

I am prepared to teach both departmental and service courses on research design, quantitative methods, statistical computing, machine learning, causal inference, and experiments. You can find copies of current and future syllabi in my website.

Service

I have devoted most of my service experience into making cutting-edge training and experience in quantitative methods as accessible as possible. Before joining Northwestern, I served as the research methods editorial assistant at the *American Political Science Review*. My primary mission in this role was to expand the reviewer pool to incorporate a higher proportion of early career scholars in the evaluation of highly technical submissions.

My long-term goal is to create or expand existing opportunities to incorporate students and early career scholars into a sustainable research pipeline centered on cutting-edge applications of statistical methods. My initial efforts in this matter have focused on working closely with undergraduate students. For example, as a graduate student at Illinois, I worked closely with an undergraduate research assistant to learn about topic modeling for text analysis together, which helped both my dissertations and the student's future career in data analysis for industry. Currently, I mentor a rising sophomore at Northwestern to incorporate them as co-authors on projects on experimental research design. With access to a bigger platform, I foresee growing toward a lab group that continuously trains and mentors students and recent graduates across all levels.