

May 2, 2024

Dear Members of the Search Committee,

I write to express my interest in your call for an Assistant Professor in Statistics. I am a postdoctoral fellow in Advanced Statistical, Causal Inference, and Computational Methodologies in the Department of Political Science at McMaster University. I received my PhD from the Department of Political Science at the University of Illinois at Urbana-Champaign under the supervision of Jake Bowers, Matt Winters, Gisela Sin, and Avital Livny.

My research uses tools from machine learning and design-based causal inference to develop standards to navigate research design tradeoffs in the social sciences. I focus on applications in which human behavior prevents us from observing the phenomena we wish to understand directly, thus demanding creativity in data analysis. My work is published, forthcoming, or under revise and resubmit in leading social science outlets discussing these topics, including the *British Journal of Political Science*, *World Development*, and the *Journal of Experimental Political Science*.

My research agenda focuses on developing tools that researchers can adopt to improve statistical precision before data collection. This is overlooked in the causal inference literature in favor of identifying unbiased estimators. Implicitly, this literature assumes that one can improve statistical precision by increasing sample size. This is not feasible in many social science applications due to resource limitations. Moreover, resource considerations aside, even the least intrusive study has an ethical mandate to identify a research design that maximizes benefits and minimizes harm at the lowest possible cost.

Focusing on randomized controlled trials, this agenda follows two strands. First, I examine cases where one can improve statistical precision without sacrificing unbiasedness, which implies unforeseen costs in other dimensions. For example, in work under review with Erin Rossiter (Notre Dame), we show how implementing alternatives to the standard experimental design, such as pre-post outcome measurement or block randomization, may attenuate the expected gains in precision via explicit or implicit sample loss. In a solo-authored piece accepted at the *Journal of Experimental Political Science*, I introduce new tools to assess the validity of estimates in double list experiments. This is a variant of the list experiment that promises more precise results but comes with under-explored questionnaire design complications. In work in progress with Inés Fynn (PUC-Chile), Verónica Pérez Bentancur (Universidad de la República), and Lucía Tiscornia (University College Dublin), we show how to combine list experiments with network scale up questions to improve the precision of prevalence rate estimates for sensitive attitudes and behaviors at the cost one additional assumption.

The second strand focuses on cases where one can improve precision by sacrificing unbiasedness deliberately. For example, in work in progress with Jake Bowers (Illinois) and Christopher Grady (USAID), we discuss the circumstances under which researchers should prefer biased yet precise estimators to analyze experimental data, including applications

to block-randomization and M-estimation. In a *SAGE Handbook* chapter with Christopher Grady and Jim Kuklinski (Illinois), we discuss the merits and challenges of increasingly complex survey experimental designs that improve precision at the cost of external validity bias.

I plan to expand my research program toward the use of machine learning to improve statistical inference. An early example comes from my dissertation work, funded by the Lemann Center for Brazilian Studies, on the effect of investigating the use of federal funds among selected mayors in Brazil on the behavior of other mayors in nearby localities. My theory suggests geographic spillovers as a quantity of interest, yet it gives no guidelines for how far away that effect would travel. This presents a tradeoff between operationalizations that introduce bias by being too narrow or too broad. I overcame this problem using a penalized regression framework to choose the optimal upper bound. This innovation received the best poster award in the 2019 Latin American Political Methodology meeting.

My teaching focuses on making applied statistics accessible to diverse audiences through a combination of flexibility and accountability. At McMaster, I teach data analysis for public policy and public opinion. At Tulane, I taught a seminar on the use of design-based causal inference to address social and political challenges in developing democracies. At Illinois, I served as a teaching assistant for statistics courses at the undergraduate and PhD levels using a flipped classroom approach. I also served as a math camp instructor for incoming graduate students for three consecutive years. I also have experience teaching political science courses in online and hybrid formats.

My involvement beyond the classroom also complements my teaching and mentoring. My work as the methods editorial assistant for the *American Political Science Review* gives me the opportunity to shape and influence the development and application of cutting-edge statistical methods in the social sciences. As a PhD student at Illinois, I started a collaborative project in which graduate students introduced their peers to new methods and organized a reading group on data science.

I am prepared to teach courses on statistical inference, statistical programming, causal inference, design of experiments, machine learning, and their application to social science and public policy. You can find copies of current and sample syllabi in my website.

I believe my expertise makes me an excellent fit at Binghamton University. If you have any questions, you can contact me via email or phone.

Sincerely,

Gustavo Diaz
Postdoctoral Fellow
Department of Political Science
McMaster University