2024 AMERICAN CAUSAL INFERENCE CONFERENCE

ABSTRACT BOOK

ORGANIZED BY:

THE SOCIETY FOR CAUSAL INFERENCE

MAY 14-17, 2024

Randomized Studies

Design-Based Causal Inference with Missing Outcomes: Missingness Mechanisms, Imputation-Assisted Randomization Tests, and Covariate Adjustment Siyu Heng* Siyu Heng, Jiawei Zhang, Yang Feng,

Design-based causal inference is immune to outcome model misspecification as its statistical validity only comes from the study design (e.g., randomization or matching design) and does not require assuming any outcome-generating distributions or models. However, design-based causal inference may still suffer from other data challenges from outcome variables, among which missingness in outcomes is a significant one. We systematically study the outcome missingness problem in design-based causal inference. First, we use the potential outcomes framework to clarify the minimal assumption (concerning the outcome missingness mechanism) needed for conducting finite-population-exact randomization tests for the null effect (i.e., Fisher's sharp null) and that needed for constructing finite-population-exact confidence sets with missing outcomes. Second, we propose a general framework called "imputation and re-imputation" for conducting finite-population-exact randomization tests in design-based causal studies with missing outcomes. Our framework can incorporate any existing outcome imputation algorithms and meanwhile guarantee finite-population-exact type-I error rate control. Third, we extend our framework to conduct covariate adjustment in an exact randomization test with missing outcomes and to construct finite-population-exact confidence sets with missing outcomes.

Randomized Studies

Near-Optimal Non-Parametric Sequential Tests and Confidence Sequences with Possibly Dependent Observations Aurelien Bibaut* Aurelien Bibaut, Nathan Kallus, Michael Lindon,

Sequential tests and their implied confidence sequences, which are valid at arbitrary stopping times, promise flexible statistical inference and on-the-fly decision making. However, strong guarantees are limited to parametric sequential tests, which suffer high type-I error rates in practice because reality isn’t parametric, or to concentration-bound-based sequences, which are overly conservative so we get wide intervals and take too long to detect effects. We consider a classic delayed-start normal-mixture sequential probability ratio test and provide the first asymptotic (in the delay) analysis under general non-parametric data generating processes. We guarantee type-I-error rates approach a user-specified α-level (primarily by leveraging a martingale strong invariance principle). Moreover, we show that the expected time-to-reject approaches the minimum possible among all α-level tests (primarily by leveraging an identity inspired by Itô’s lemma). Together, our results establish these (ostensibly parametric) tests as general-purpose, non-parametric, and near-optimal. We illustrate this via numerical experiments and a retrospective re-analysis of A/B tests at Netflix.

Randomized Studies

Identification and estimation of causal effects using non-concurrent controls in platform trials Michele Santacatterina* Michele Santacatterina, Federico Macchiavelli, Xinyi Zhang, Ivan Diaz,

Platform trials are multi-arm designs that simultaneously evaluate multiple treatments for a single disease within the same overall trial structure. Unlike traditional randomized controlled trials, they allow treatment arms to enter and exit the trial at distinct times while employing a shared control arm. In platform trials, concurrent controls join alongside new treatments, while non-concurrent controls are already enrolled, potentially introducing time-related biases. The central question revolves around the effective utilization of non-concurrent controls to estimate treatment effects in platform trials. Specifically, what estimands should be used to evaluate the causal effect of a treatment versus a shared control? Under what assumptions can these estimands be identified and estimated? Do we achieve any efficiency gains? In this project, we use structural causal models and counterfactuals to clarify estimands and formalize their identification in the presence of non-concurrent controls in platform trials. We also provide estimators based on outcome regression, inverse probability weighting, and doubly robust estimators for their estimation. Additionally, we discuss efficiency gains, demonstrate their performance in a simulation study, and apply them using data from the Adaptive COVID-19 Treatment Trial (ACTT) platform trial.

Randomized Studies

Debiased regression adjustment in completely randomized experiments with moderately high-dimensional covariates Yuhao Wang* Yuhao Wang, Xin Lu, Fan Yang,

Completely randomized experiment is the gold standard for causal inference. When the covariate information for each experimental candidate is available, one typical way is to include them in covariate adjustments for more accurate treatment effect estimation. In this paper, we investigate this problem under the randomization-based framework, i.e., that the covariates and potential outcomes of all experimental candidates are assumed as deterministic quantities and the randomness comes solely from the treatment assignment mechanism. Under this framework, to achieve asymptotically valid inference, existing estimators usually require either (i) that the dimension of covariates p grows at a rate no faster than O(n^{2 / 3}); or (ii) certain sparsity constraints on the linear representations of potential outcomes with respect to possibly high-dimensional covariates. In this paper, we consider the moderately high-dimensional regime where p is allowed to be in the same order of magnitude as n. We develop a novel debiased estimator with a corresponding inference procedure and establish its asymptotic normality under mild assumptions. Our estimator is model-free and does not require any sparsity constraint on potential outcome's linear representations. We also discuss its asymptotic efficiency improvements over the unadjusted treatment effect estimator under different dimensionality constraints. Numerical analysis confirms that our estimator performs well in moderately high dimensions.

Randomized Studies

Adaptive Neyman Allocation Jinglong Zhao* Jinglong Zhao,

In experimental design, Neyman allocation refers to the practice of allocating subjects into treated and control groups, potentially in unequal numbers proportional to their respective standard deviations, with the objective of minimizing the variance of the treatment effect estimator. This widely recognized approach increases statistical power in scenarios where the treated and control groups have different standard deviations, as is often the case in social experiments, clinical trials, marketing research, and online A/B testing. However, Neyman allocation cannot be implemented unless the standard deviations are known in advance. Fortunately, the multi-stage nature of the aforementioned applications allows the use of earlier stage observations to estimate the standard deviations, which further guide allocation decisions in later stages. In this paper, we introduce a competitive analysis framework to study this multi-stage experimental design problem. We propose a simple adaptive Neyman allocation algorithm, which almost matches the information-theoretic limit of conducting experiments. Using online A/B testing data from a social media site, we demonstrate the effectiveness of our adaptive Neyman allocation algorithm, highlighting its practicality especially when applied with only a limited number of stages.

Randomized Studies

Designing target trial emulations for COVID-19 pharmacotherapy effectiveness studies – challenges and implications of assigning time zero David Bui* David Bui, Kristina Bajema, Kristin Berry, Lei Yan, Yuan Huang, Yuli Li, Nallakkandi Rajeevan, Stephanie Argraves, Valerie Smith, Matthew Maciejewski, Amy Bonhert, Denise Hynes, Mihaela Aslan, George Ioannou,

Real-world pharmacotherapy effectiveness (PE) studies should be carefully designed using target trial emulation (TTE) principles for valid causal inference. Designing target trials of outpatient COVID-19 PE presents unique challenges, especially in determining time zero—the point in a trial at which a patient is determined eligible, randomized to a treatment group, and starts follow-up. In TTEs, defining time zero can be difficult when only one group, namely treated persons, has a clear time zero because there is no obvious time zero for comparators not initiating an alternative treatment. There are several approaches for defining time zero (index date) in observational studies of COVID-19 pharmacotherapies: 1) using the date of positive SARS-CoV-2 test in both treated and untreated persons, 2) using the date of treatment initiation in treated persons and test-positive date in untreated persons, or 3) using the date of treatment initiation in treated persons and index date corresponding to the same number of days following the test-positive date in untreated persons. We describe how well each of these approaches emulates a randomized trial of oral antiviral COVID-19 treatment versus no treatment. Using real-world data from the U.S. Veterans Health Administration, we conduct a set of TTEs using each time zero approach, describe relative strengths, implications, and limitations for causal inference.

Randomized Studies

Statistical challenges of pragmatic randomized trials with intervention-dependent outcome assessment processes Jennifer Bobb* Jennifer Bobb, Sungtaek Son, Melissa Anderson, Lynn DeBar, Katharine Bradley,

Pragmatic trials are increasingly being conducted that use real-world data such as electronic health records (EHRs) to define study outcomes. Unlike traditional trials in which outcomes are collected at pre-specified time points, follow-up times from EHRs are irregularly spaced, not controlled by the study team, and may be outcome dependent. In this talk, we explore issues that arise when interventions being studied affect outcome assessment, including frequency and timing of follow-up measures. This work is motivated by the MICARE trial among primary care patients with opioid use disorder and depression, in which the intervention is expected to increase documentation of the depression outcome in the EHR (Patient Health Questionnaire [PHQ] measure of depression). We conducted simulations to examine the performance of common statistical approaches in the setting of intervention-dependent outcome assessment times, including simple approaches that select a single follow-up measure per person (e.g., score closest to 12 months), and longitudinal approaches that use all follow-up data. We additionally clarify which estimands are being estimated when observation times are intervention dependent, including time-point specific and (weighted) time-averaged treatment effects. Our results informed the selection of statistical approach for the MICARE trial and have important implications for the design of trials with intervention-dependent measurement processes.

Generalizability/Transportability

Data Fusion under Disparate Outcome Measures Harsh Parikh* Harsh Parikh, Elizabeth Stuart, Kara Rudolph,

Randomized controlled trials (RCTs) serve as the cornerstone for understanding causal effects, yet extending inferences to target populations presents challenges due to effect heterogeneity and underrepresentation. Discrepancies in distributions across individual characteristics between trial and target populations often impair treatment effect extrapolations, especially for underrepresented groups, leading to inaccuracies in decision-making. Our paper addresses the critical issue of identifying and managing underrepresented subgroups in RCTs, proposing a novel framework for refining target populations to enhance treatment effect generalizability. We introduce an optimization-based approach, Rashomon Set of Optimal Trees (ROOT), to characterize underrepresented groups. By minimizing the variance of the target average treatment effect (TATE) estimate, ROOT optimizes the target subpopulation distribution, ensuring more precise treatment effect estimations. Notably, ROOT generates interpretable characteristics of the underrepresented population, aiding researchers in communicating for targeted recruitment in future trials. Through extensive evaluation using synthetic data experiments with complex structures, our approach demonstrates superior precision enhancement and interpretability compared to alternatives. Applying our methodology to an RCT on medication for Opioid Use Disorder (MOUD), we investigate discrepancies between trial results and real-world effectiveness.

Randomized Studies

Selective randomization inference for adaptive studies Tobias Freidling* Tobias Freidling, Zijun Gao, Qingyuan Zhao,

Many clinical trials follow a design with multiple stages: After each stage, the data is provisionally analysed and – based on these results – the recruitment of participants for the next stage as well as the administered treatment is chosen adaptively. For instance, we may want to exclude poorly performing drugs early or gather more samples from a certain subpopulation that shows a potentially beneficial response.
Analysing such adaptive studies is challenging as the data is used twice: (1) for selection of the design of later stages and the null hypothesis, (2) for testing the null hypothesis with the data generated under the chosen design. Since the data generating mechanism and null hypothesis are not pre-specified, classical statistical methods do not provide valid inference.

Existing solutions are often limited in scope and usually specific to a certain design; in this work, we propose a general framework that can handle all kinds of designs and adaptive choices. Our approach uses concepts from the post-selection inference literature to develop a selective randomization p-value. Noteably, we do not require any assumptions on the law of the outcomes and covariates or on the dependence structure between different participants. We show that our method improves power compared to other valid tests while still controlling the selective type-I error. Moreover, we construct confidence intervals and discuss different methods to compute the selective randomization p-value.

Causal Inference in Networks

Causal clustering: design of cluster experiments under network interference Lihua Lei*
Lihua Lei, Davide Viviano, Guido Imbens, Brian Karrer, Okke Schrijvers, Liang Shi,

This paper studies the design of cluster experiments to estimate the global treatment effect in the presence of spillovers on a single network. We provide an statistical decision-theoretic framework to choose the clustering that minimizes the worst-case mean-squared error of the estimated global treatment effect. We show that the optimal clustering can be approximated as the solution of a novel penalized min-cut optimization problem computed via off-the-shelf semi-definite programming algorithms. Our analysis also characterizes easy-to-check conditions to choose between a cluster or individual-level randomization. We illustrate the method’s properties using unique network data from the universe of Facebook’s users and existing network data from a field experiment.

Instrumental Variables

DRIVE: A Distributionally Robust Instrumental Variable Estimation Framework Zhaonan Qu* Zhaonan Qu, Yongchan Kwon,

We propose a distributionally robust formulation of the classical instrumental variable (IV) estimation framework, motivated by common challenges to IV estimators in practice, such as invalid and weak instruments and poor finite sample properties. When the distributional uncertainty set is a Wasserstein ball, the resulting estimator, which we call Wasserstein DRIVE, is consistent whenever the robustness parameter is bounded above by the largest singular value of the first stage coefficient and instruments are valid. We propose two data-driven procedures to select the penalty parameter, based on the first stage regression coefficient and the score quantile estimated using a nonparametric bootstrap. Thanks to its regularization and robustness properties, Wasserstein DRIVE could be preferable in practice, particularly when model assumptions are potentially violated. Simulation studies confirm the superior finite sample performance of Wasserstein DRIVE with valid and invalid instruments. Application to the Card dataset on educational returns also demonstrates its robustness at prediction under covariate shifts.

Instrumental Variables

Nested Instrumental Variables Design: Switcher Average Treatment Effect, Identification, Efficient Estimation and Generalizability Rui Wang* Rui Wang, Yingqi Zhao, Oliver Dukes, Bo Zhang,

Instrumental variables (IV) are a commonly used tool in estimating causal effect from non-randomized data. A prototype of an IV is a randomized trial with no compliance where the randomized treatment assignment serves as an IV for the nonignorable treatment uptake. Under a monotonicity assumption, a valid IV nonparametrically identifies the average treatment effect among a non-identifiable complier subgroup, whose generalizability is often under much debate. In many studies, there could exist multiple versions of an IV, for instance, different nudges to take the treatment in different study sites in a clinical trial. These different versions of an IV may result in different compliance rates and offer a unique opportunity to study IV estimates' generalizability. In this article, we introduce a novel nested IV assumption and study the identification of the average treatment effect among two latent subgroups: always-compliers and switchers, who are defined based on the joint potential treatment uptake under two versions of a binary IV. We derive the conanical gradient for the SWitcher Average Treatment Effect (SWATE) and propose efficient estimator. We then propose formal statistical tests of the principal ignorability assumption based on comparing the conditional average treatment effect among the always-compliers and that among the switchers under the nested IV framework. We apply our method to the Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial.

Instrumental Variables

Estimation of the Number Needed to Treat, the Number Needed to be Exposed, and the Exposure Impact Number with Instrumental Variables Valentin Vancak* Valentin Vancak, Arvid Sjölander,

The Number needed to treat (NNT) is an efficacy index defined as the average number of patients needed to treat to attain one additional treatment benefit. In observational studies, specifically in epidemiology, the adequacy of the populationwise NNT is questionable since the exposed group characteristics may substantially differ from the unexposed. To address this issue, groupwise efficacy indices were defined: the Exposure Impact Number (EIN) for the exposed group and the Number Needed to be Exposed (NNE) for the unexposed. Each defined index answers a unique research question since it targets a unique sub-population. In observational studies, the group allocation is typically affected by confounders that might be unmeasured. The available estimation methods that rely either on randomization or the sufficiency of the measured covariates for confounding control will result in inconsistent estimators of the true NNT (EIN, NNE) in such settings. Using Rubin’s potential outcomes framework, we explicitly define the NNT and its derived indices as causal contrasts. Next, we introduce a novel method that uses instrumental variables to estimate the three aforementioned indices in observational studies. Finally, we present two analytical examples and a corresponding simulation study. The simulation study illustrates that the novel estimators are consistent, unlike the previously available methods, and their confidence intervals meet the nominal coverage rates.

Instrumental Variables

Debiased Multivariable Mendelian Randomization Yinxiang Wu* Yinxiang Wu, Hyunseung Kang, Ting Ye,

Multivariable Mendelian randomization (MVMR) uses genetic variants as instrumental variables to infer the direct effect of multiple exposures on an outcome. Compared to univariable Mendelian randomization, MVMR is less prone to horizontal pleiotropy and enables estimation of the direct effect of each exposure on the outcome. However, MVMR faces greater challenges with weak instruments–genetic variants that are weakly associated with some exposures conditional on the other exposures. This article focuses on MVMR using summary data from genome-wide association studies (GWAS). We provide a new asymptotic regime to analyze MVMR estimators with many weak instruments, allowing for linear combinations of exposures to have different degrees of instrument strength, and formally show that the popular multivariable inverse-variance weighted (MV-IVW) estimator's asymptotic behavior is highly sensitive to instruments' strength. We then propose a multivariable debiased IVW (MV-dIVW) estimator which effectively reduces the asymptotic bias from weak instruments in MV-IVW, and introduce an adjusted version, MV-adIVW, to improve MV-dIVW's finite-sample robustness. We establish the theoretical properties of our proposed estimators and extend them to handle balanced horizontal pleiotropy. We conclude by demonstrating the performance of our proposed methods in simulated and real datasets. We implement this method in the R package mr.divw.

Machine Learning and Causal Inference

Double Descent in Double Machine Learning Jann Spiess* Jann Spiess, Guido Imbens, Amar Venugopal,

Motivated by a literature on the double-descent phenomenon in machine learning, we consider highly over-parameterized models in causal estimation using double machine learning. We build upon our earlier work on double descent in causal imputation (https://arxiv.org/abs/2305.00700; NeurIPS 2023) to investigate high-dimensional linear regression for estimating outcome regressions and inverse propensity score weights in doubly-robust causal estimators. Specifically, we consider linear-regression models with many more covariates than sample size. In such models, there may be so many free parameters that the model interpolates the training data perfectly. For outcome models in doubly-robust estimation, we find that such interpolating regressions can outperform simple ones, provided that we use appropriate sample splitting. For inverse propensity score weighting, we instead find that estimation strategies that plug in estimated propensity scores from overparameterized models fail dramatically. As a potential remedy, we instead investigate the direct estimation of balancing weights using the automated estimation of Riesz representers with interpolating linear models.

Bayesian Causal Inference

Bayesian Causal Models from a Weighting Perspective: Balance, Bias, and Double Robustness Jared Murray* Jared Murray, Avi Feller,

Many popular Bayesian regression models for causal inference produce estimates that are linear in outcomes and can be cast as weighting estimators. Examples include Bayesian Additive Regression Trees (BART), Bayesian Causal Forests (BCF), and Gaussian process regression (including regression with conditionally Gaussian priors). We consider estimating average effects in generic target populations under ignorable selection into treatment.

We show that the model-implied weights have a dual interpretation as regularized minimax balancing weights or, equivalently, estimates of a Riesz representer of the estimand (e.g., inverse propensity weights when estimating the PATE). We derive the bias of the posterior mean when the outcome model is partially correct, describe how to minimize it via model and prior specification, and examine the role of the implied estimate of the Riesz representer under misspecification.

Finally, we construct doubly robust estimates via Bayesian decision theory using a new loss function that prioritizes bias minimization. Minimizing the posterior expected loss leads to doubly robust augmented weighting estimators without modifying the prior or likelihood. These Bayes estimates belong to the family of “automatically debiased” machine learning estimates of causal effects, with tuning parameters that can be informed by data through Bayesian updating. We conclude by discussing the construction of Frequentist and (quasi-)Bayesian uncertainty intervals.

Machine Learning and Causal Inference

Model-Agnostic Covariate-Assisted Inference on Partially Identified Causal Effects Asher Spector* Asher Spector, Lihua Lei, Wenlong Ji,

Many causal estimands are only partially identifiable since they depend on the unobservable joint distribution between potential outcomes. Stratification on pretreatment covariates can yield sharper partial identification bounds; however, unless the covariates are discrete with relatively small support, this approach typically requires consistent estimation of the conditional distributions of the potential outcomes given the covariates. Thus, existing approaches may fail under model misspecification or if consistency assumptions are violated. In this study, we propose a unified and model-agnostic inferential approach for a wide class of partially identified estimands, based on duality theory for optimal transport problems. In randomized experiments, our approach can wrap around any estimates of the conditional distributions and provide uniformly valid inference, even if the initial estimates are arbitrarily inaccurate. Also, our approach is doubly robust in observational studies. Notably, this property allows analysts to use the multiplier bootstrap to select covariates and models without sacrificing validity even if the true model is not included. Furthermore, if the conditional distributions are estimated at semiparametric rates, our approach matches the performance of an oracle with perfect knowledge of the outcome model. Finally, we propose an efficient computational framework, enabling implementation on many practical problems in causal inference.

Machine Learning and Causal Inference

Nonparametric Identification and Estimation of Average Treatment Effects with a Latent Exposure Ying Zhou* Ying Zhou, Eric Tchetgen Tchetgen,

In many practical scenarios, the direct observation of exposure variables is not feasible, hindering the ability to draw valid causal inferences. To address this issue, we propose a method that employs two proxy variables of a binary latent exposure, enabling the identification of the average treatment effect (ATE). We then derive the efficient influence function for the ATE, and construct an efficient nonparametric estimator. A significant obstacle to our approach is the estimation of nuisance functions involving the latent exposure, which prevents the direct application of standard machine learning algorithms. To resolve this, we introduce a novel EM-like algorithm, thus adding a practical dimension to our theoretical contributions. This methodology can be adapted to analyze causal functionals beyond ATE, provided two proxies of the latent exposure are available.

Machine Learning and Causal Inference

Learning Causal Representations from General Environments: Identifiability and Intrinsic Ambiguity JIKAI JIN* JIKAI JIN, Vasilis Syrgkanis,

This paper studies causal representation learning, the task of recovering high-level latent variables and their causal relationships from low-level data that we observe, assuming access to observations generated from multiple environments. While existing works are able to prove full identifiability of the underlying data generating process, they typically assume access to single-node, hard interventions which is rather unrealistic in practice. The main contribution of this paper is characterize a notion of identifiability which is provably the best one can achieve when hard interventions are not available. First, for linear causal models, we provide identifiability guarantee for data observed from general environments without assuming any similarities between them. While the causal graph is shown to be fully recovered, the latent variables are only identified up to an effect-domination ambiguity (EDA). We then propose an algorithm, LiNGCReL which is guaranteed to recover the ground-truth model up to EDA, and we demonstrate its effectiveness via numerical experiments. Moving on to general non-parametric causal models, we prove the same idenfifiability guarantee assuming access to groups of soft interventions. Finally, we provide counterparts of our identifiability results, indicating that EDA is basically inevitable in our setting.

Difference in Differences

Practical Guidance on Whether and When to Aggregate Individual-Level Data for Causal Health Policy Evaluation Nicholas Seewald* Nicholas Seewald, Kayla Tormohlen, Beth McGinty, Elizabeth Stuart,

Health policy researchers often have questions about the effects of state policy on individual-level outcomes collected over multiple time periods. Such questions might be addressed using, for example, a large health insurance claims database that tracks individuals' receipt of a particular treatment. An open question is whether the researcher can or should "roll-up" (i.e., aggregate, average, etc.) this individual-level data to the policy level when assessing the effects of state policy. Rolling up the data offers a clear computational advantage since it makes the individual-level big data question much smaller. However, existing literature does not sufficiently address whether and when aggregation is disadvantageous due to loss of individual-level information. Here, we examine the statistical performance of difference-in-differences approaches that permit the use of either individual- or aggregate-level data to offer practical guidance on whether and when to roll up. Our guidance is based on simulation models which allow us to make fair comparisons between analytic methods under a variety of controlled conditions. We also discuss our recommendations in the context of a study designed to assess the effects of state medical cannabis laws on opioid prescribing among patients with chronic non-cancer pain.

Difference in Differences

Robust Transportation of Public Policy Effects to New Contexts Gary Hettinger* Gary Hettinger, Youjin Lee, Nandita Mitra,

Public policies are crucial tools for numerous societal goals like improving health and economic outcomes. As a result, there have been substantial recent methodological developments for evaluating policy effects in regions where such policies are implemented. However, policies often have largely different consequences when implemented in new regions that may differ in their demographic and geographic characteristics as well as their implementation strategy. These differences limit the use of traditional methods and analyses for future decision-making.

To improve upon these limitations, we develop methodology to transport policy effects from an implementing region to a new region that may be considering a similar policy. Our methods build upon classical difference-in-differences approaches and semiparametric theory to first flexibly estimate effect heterogeneity and then adapt these findings to the nuances of a new region. In addition to semiparametric estimators, we also present a new set of identification assumptions necessary for valid transportation and tools for assessing the validity of potential projections. We apply our methods to study effects of excise tax policies in the US.

Difference in Differences

Identifying and estimating causal effects using difference-in-differences under network dependency and interference Michael Jetsupphasuk* Michael Jetsupphasuk, Didong Li, Michael Hudgens,

Differences-in-differences (DiD) is a causal inference framework for observational panel data that allows for unmeasured confounding but assumes parallel outcome trajectories among treatment groups under the (possible) counterfactual of receiving a specific treatment. We study DiD under network dependency and interference, where outcomes may be correlated and treatments assigned to a unit may affect outcomes in neighboring units. The proposed methods accommodate general interference provided there exists a known exposure mapping that summarizes treatments in interfering units. This framework includes the bipartite setting where treatment and outcome units are different. The main estimand of interest generalizes recently proposed estimands and is a time-varying analogue of the average treatment effect among the treated where potential outcomes may depend on multi-valued or continuous exposure histories. We identify the causal estimand under parallel trends and propose outcome regression, inverse probability weighted, and doubly robust estimators. The methods are evaluated in simulations and applied to study the effects of adopting emission control technologies in coal power plants on county-level mortality due to cardiovascular disease.

Difference in Differences

Estimating Counterfactual Matrix Means with Short Panel Data Brad Ross* Brad Ross, Lihua Lei,

We develop a more flexible approach for identifying and estimating average counterfactual outcomes when several but not all possible outcomes are observed for each unit in a large cross section. Such settings include event studies and studies of outcomes of “matches” between agents of two types, e.g. workers and firms or people and places. When outcomes are generated by a factor model that allows for low-dimensional unobserved confounders, our method yields consistent, asymptotically normal estimates of counterfactual outcome means under asymptotics that fix the number of outcomes as the cross section grows and general outcome missingness patterns, including those not accommodated by existing methods. Our method is also computationally efficient, requiring only a single eigendecomposition of a particular aggregation of any factor estimates constructed using subsets of units with the same observed outcomes. In a semi-synthetic simulation study based on matched employer-employee data, our method performs favorably compared to a Two-Way-Fixed-Effects-model-based estimator.

The paper draft can be found here: https://arxiv.org/abs/2312.07520

Difference in Differences

Evaluating causal impact of device deprecation and beyond Can Cui* Cindy Cui,

Older devices present security risk and high maintenance costs as technology evolves and new consumer devices get adopted at faster pace. Business generally approach deprecation their services on certain devices in a planned manner cautiously, balancing potential impact on device manufacturers and consumers. We study the impact of depreciating older devices using observational causal method. We covers different use cases, including simple one-time treatment with valid control group where differences-in-differences or synthetic control works well, and more complex cases where we lack control group or the timing and implementation of treatment is more complicated. In addition to examining performance of these methods, we also build a predicted impact model to support ongoing and future business decisions on which device to retire and when to do so, in order to minimize negative impact on customer and device partners while reducing engineering maintenance costs and security risks. We will share our learnings on performance of various observational causal methods in this setting, and our efforts to setup semi-automatic process to apply most suitable approach for impact evaluation. Furthermore, we will also discuss how we can go beyond measurement and projecting forward potential impact to guide future interventions.

Bayesian Causal Inference

Data Fusion for Heterogeneous Treatment Effect Estimation with Multi-Task Gaussian Processes Evangelos Dimitriou* Evangelos Dimitriou, Edwin Fong, Karla Diaz-Ordaz, Brieuc Lehmann,

In the pursuit of reliable causal predictions, balancing internal and external validity is crucial. Randomised Controlled Trials (RCTs), favoured for their internal validity due to randomisation, often encounter challenges in generalising findings due to strict eligibility criteria. Observational studies conversely, provide external validity advantages through larger sample sizes but compromise internal validity due to unmeasured confounding. Integrating RCTs with observational studies is a promising way to get the best of both worlds. In this context, we propose a novel Bayesian nonparametric approach leveraging multi-task Gaussian Processes to seamlessly integrate findings from both RCTs and observational studies. Our method treats potential outcomes from distinct data sources as tasks in a multi-task learning framework, modelling their relationships through a shared covariance function. This integration enables unbiased estimations within the target population, addressing both internal and external validity concerns. Our approach outperforms cutting-edge methods in point predictions across the covariate support of the target population. Additionally, it provides a quantifiable measure of uncertainty, offering a comprehensive understanding of the reliability of the estimated treatment effects. We demonstrate the performance of our approach through multiple simulation studies and a real world data application, showcasing its robust performance in diverse scenarios.

Bayesian Causal Inference

Causally Sound Priors for Binary Experiments Nicholas Irons* Nicholas Irons, Carlos Cinelli,

We introduce the BREASE framework for the Bayesian analysis of randomized controlled trials with a binary treatment and a binary outcome. Approaching the problem from a causal inference perspective, we propose parameterizing the likelihood in terms of the baseline risk, efficacy, and side effects of the treatment, along with a flexible, yet intuitive and tractable jointly independent beta prior distribution on these parameters, which we show to be a generalization of the Dirichlet prior for the joint distribution of potential outcomes. Our approach has a number of desirable characteristics when compared to current mainstream alternatives: (i) it naturally induces prior dependence between expected outcomes in the treatment and control groups; (ii) as the baseline risk, efficacy and side effects are quantities inherently familiar to clinicians, the hyperparameters of the prior are directly interpretable, thus facilitating the elicitation of prior knowledge and sensitivity analysis; and (iii) it admits analytical formulae for the marginal likelihood, Bayes factor, and other posterior quantities, as well as exact posterior sampling via simulation, in cases where traditional MCMC fails. Empirical examples demonstrate the utility of our methods for estimation, hypothesis testing, and sensitivity analysis of treatment effects.

Bayesian Causal Inference

Identified vaccine efficacy for binary post-infection outcomes under misclassification without monotonicity Rob Trangucci* Rob Trangucci, Yang Chen, Jon Zelner,

In order to meet regulatory approval, pharmaceutical companies often demonstrate that new vaccines reduce the total risk of a post-infection outcome like transmission, symptomatic illness, or death in randomized, placebo-controlled trials.
One can use principal stratification to partition this causal effect into vaccine efficacy against infection, and the principal effect of vaccine efficacy on post-infection outcomes in always-infected patients.
Unfortunately, even under strong assumptions, these principal effects are generally unidentifiable.
We develop a novel method to nonparametrically point identify these principal effects while eliminating the typical monotonicity assumption and allowing for measurement error in both infection and post-infection outcomes.
Furthermore, we show that our results readily extend to multiple treatments.
Our method takes advantage of the geographic heterogeneity of disease incidence, and well-measured biologically-relevant categorical pretreatment covariates, each a feature of many vaccine trials.
We show that a Bayesian version of our method can be applied to a variety of clinical trial settings where vaccine efficacy against infection and a post-infection outcome can be jointly inferred, and investigate the sensitivity to prior specification using simulation studies.
Finally, we apply this method to an influenza vaccine trial to yield new insights into vaccine efficacy against symptomatic illness.

Bayesian Causal Inference

Estimating the Returns from an Experimentation Program Simon Ejdemyr* Simon Ejdemyr, Martin Tingley, Yian Shang, Travis Brooks,

We describe the development, validation and implementation of a Bayesian hierarchical model used by Netflix for estimating the returns of innovation areas that leverage A/B testing. The model provides a trusted source for estimates of the cumulative returns from experiment launches, and integration into Netflix's flagship experimentation UI has facilitated a better understanding of how different testing areas improve business outcomes at different rates. This understanding can help company leaders prioritize the most promising innovation programs, or pivot innovation strategies in areas that show diminishing returns. In fact, surfacing these views at Netflix has already streamlined previously time-consuming annual review processes and goal tracking.

The primary statistical challenge the model addresses is overestimation of cumulative treatment effects, due to selection on winners. This is accomplished via hierarchical shrinkage. The model's first level facilitates information borrowing across tests within the same testing area, while the second level models the distribution of within-test effects. The model imposes weak parametric structure on the assumed distribution of the true treatment effects, allowing for Gaussian or fat-tailed structures. We show that this model validates well against holdback tests (large retests of launched treatments): across two critical testing areas at Netflix, the model removed upward bias, consistent with the holdbacks.

Synthetic Control Method

Estimating Policy Effects using Lagged Outcome Values to Impute Counterfactuals Beth Ann Griffin* Beth Ann Griffin, David Powell, Tal Wolfson,

Causal inference requires estimation of counterfactuals. With panel data, it is common to use information from the pre-treatment period of the treated units combined with information from untreated units to impute these counterfactual outcomes. Difference-in-differences methods, such as two-way fixed effects models, have frequently been used to predict counterfactual outcomes assuming additive fixed unit and time effects. Despite the growth of panel data estimators which impute counterfactuals, little research considers using lagged outcomes to directly predict post-treatment counterfactual outcomes. We discuss implementation of a “lagged outcome model (LOM)”. The LOM estimator involves regressing the outcome variable on pre-treatment period lags using untreated units only, penalizing the inclusion of additional lags. Given the resulting estimates, it is then possible to impute the counterfactual for the treated observations. We compare the LOM approach to commonly used methods in applied work including the synthetic control method, synthetic difference-in-differences, synthetic control estimation with bias-correction, de-meaned synthetic control estimation, the matrix completion method, and a penalized vertical regression. The LOM performs well relative to these estimators in simulations regardless of pre-period length. These results suggest that LOM should become a more standard part of the panel data imputation toolkit for empirical researchers.

Synthetic Control Method

History versus Unobservable Confounding in Causal Inference with Panel Data: A Design-based Perspective Ye Wang* Ye Wang, Yiqing Xu,

Should researchers control for past values of variables or unobservable factors such as the fixed effects when attempting to establish causality in panel data, or is it possible to account for both? The paper investigates this question from a design-based perspective, which distinguishes identification assumptions on the treatment assignment process from structural restrictions for estimators to yield interpretable results. We propose a framework to unify two identification assumptions that are commonly invoked to justify the choice, sequential ignorability and strict exogeneity, and review available methods under each. We argue that fixed effects models are compatible with the presence of dynamic effects when treatment status does not reverse for any unit. In such scenarios, many existing methods automatically account for both types of confounders and generate estimates that agree with each other. Otherwise, structural restrictions on the persistence or heterogeneity of dynamic effects must be imposed for researchers to rely on fixed effects models to address the challenge of causal identification. We provide guidance for applied researchers to choose from potential options under various circumstances and propose a novel estimator when a long pre-treatment period exists. We substantiate these propositions through Monte Carlo simulations and a case study examining the effect of democracy on economic development.

Event Studies

Anatomy of Two-Way Fixed Effects Models: Experimental Design Principles, Use of Information, and Robust Estimation Zhu Shen* Zhu Shen, Yuzhou Lin, Ambarish Chattopadhyay, Jose Zubizarreta,

In recent decades, event studies have emerged as a leading methodology in health and social research for evaluating the causal effects of discrete interventions. In this paper, we provide a novel characterization of the classical dynamic two-way fixed effects (TWFE) regression estimator for event studies. Our decomposition is expressed in closed-form and reveals, in finite samples and without approximations, the hypothetical experiment that TWFE regression adjustments approximate. This decomposition offers insights into how standard regression estimators use information from various units and time points, generalizing the notion of forbidden comparison noted in the literature in simpler settings. We propose a robust weighting approach for estimation in event studies, which allows investigators to progressively build larger valid weighted contrasts by leveraging, in a sequential manner, increasingly stronger assumptions on the potential outcomes and the assignment mechanism. We provide weighting diagnostics and visualization tools. We illustrate these methods in a case study of the impact of divorce reforms on female suicide.

Difference in Differences

Marginal Structural Nested Mean Models Under Parallel Trends Zach Shahn* Zach Shahn, Oliver Dukes, James Robins, Andrea Rotnitzky,

Suppose an investigator conducting a DiD analysis is interested in modeling effect heterogeneity as a function of only a subset of the adjustment variables required to satisfy conditional parallel trends. Abadie (2005) proposed an estimator for the parameter of such a marginal effect heterogeneity model in the point exposure setting. Using marginal Structural Nested Mean Models (mSNMMs), we provide a doubly robust and efficient estimator of this parameter under the same assumptions, and also generalize the problem to the time-varying setting. We thus enable DiD practitioners to model heterogeneity of time-varying treatment effects as a function of a low dimensional time-varying covariate even when the adjustment set is large. For example, we use mSNMMs to estimate how effects of Medicaid expansion vary with eligibility thresholds at time of expansion, while adjusting for demographic variables. Modeling lower dimensional effect modification reduces the risk of serious model misspecification and can improve interpretability. In the special case when the effect modifiers of interest are the empty set, mSNMMs target similar estimands to the host of other recently developed and widely used time-varying DiD methods, but with certain advantages. For example, under conditional parallel trends assumptions for dynamic regimes, mSNMMs can identify effects of sustained treatment regimes even when treatment switches on and off in the data, a source of consternation in the DiD literature.

Synthetic Control Method

Augmented Balancing for Difference-in-Differences: A Synthesis Apoorva Lal* Apoorva Lal, Yiqing Xu, Ziyi Liu,

We synthesize recent estimators in synthetic control literature and show that these estimators are special cases of augmented balancing in a (potentially staggered) difference-in-differences setup, differing in their choice of outcome model and the balancing scheme generating the weights. Through a series of Monte Carlo exercises, we demonstrate that, unlike other estimators that are typically fragile in certain scenarios, the synthetic difference-in-differences (SDID) estimator is remarkably robust to different data-generating processes, primarily due to improved overlap (both cross-sectional and temporal) generated by dual weights and its ability to accommodate weak signals in low-rank structures. Moreover, we show that proper data pre-processing, coupled with a flexible outcome model such as a factor model incorporating covariates, can improve the performance of SDID. We offer practical recommendations and a software implementation for practitioners.

Causal Inference in Networks

Causal Inference when Intervention Units and Outcome Units Differ Fabrizia Mealli* Fabrizia Mealli, Georgia Papadogeorgou, Guido Imbens,

We study inference for causal effects in settings characterized by interference stemming from having two distinct sets of units: units to which the intervention is applied and units on which the outcomes are measured. We call this bipartite interference: treatments applied to one intervention unit can affect multiple outcome units, and the outcome of a unit may depend on the treatments applied to multiple intervention units.
Examples of this setting can be found across many disciplines. In air pollution epidemiology, the interventional units could be pollution emitters such as power plants (intervention units) which may install a filter on their smokestack or not, and the outcome units could be members of the population residing within specific geographical areas. Similarly, in the economics of housing, housing prices at different locations (outcome units) may be affected by whether appropriate cleaning has taken place in nearby contaminated hazardous-waste disposal sites (intervention units).
We consider both usual and new causal estimands, highlithing similarities and differences with more common settings of causal inference with unit-to-unit interference. Estimators for these quantities in bipartite settings will be introduced and their performance evaluated from a design-based perspective. For inference, finite sample and asymptotic properties will be investigated. Optimal designs for such effects will be discussed that exploit the topology of the bipartite graph.

Machine Learning and Causal Inference

Targeted Machine Learning for Average Causal Effect Estimation Using the Front-Door Functional Anna Guo* Anna Guo, Razieh Nabi, David Benkeser,

Evaluating the average causal effect (ACE) of a treatment on an outcome often involves overcoming the challenges posed by confounding factors in observational studies. A traditional approach uses the back-door criterion, seeking adjustment sets to block confounding paths between treatment and outcome. However, this method struggles with unmeasured confounders. As an alternative, the front-door criterion offers a solution, even in the presence of unmeasured confounders between treatment and outcome. This method relies on identifying mediators that are not directly affected by these confounders and that completely mediate the treatment's effect. Here, we introduce novel estimation strategies for the front-door criterion based on the targeted minimum loss-based estimation theory. Our estimators work across diverse scenarios, handling binary, continuous, and multivariate mediators. They leverage data-adaptive machine learning algorithms, minimizing assumptions and ensuring key statistical properties like asymptotic linearity, double-robustness, efficiency, and valid estimates within the target parameter space. We establish conditions under which the nuisance functional estimations ensure the root n-consistency of ACE estimators. Our numerical experiments show the favorable finite sample performance of the proposed estimators. We demonstrate the applicability of these estimators to analyze the effect of early stage academic performance on future yearly income

Machine Learning and Causal Inference

Data-Driven Influence Functions for Optimization-Based Causal Inference Angela Zhou* Angela Zhou,

We study a constructive algorithm that approximates Gateaux derivatives for statistical functionals by finite differencing, with a focus on functionals that arise in causal inference. We study the case where probability distributions are not known textit{a priori} but need to be estimated from data. These estimated distributions lead to empirical Gateaux derivatives, and we study the relationships between empirical, numerical, and analytical Gateaux derivatives. Starting with a case study of the interventional mean (average potential outcome), we delineate the relationship between finite differences and the analytical Gateaux derivative. We then derive requirements on the rates of numerical approximation in perturbation and smoothing that preserve the statistical benefits of one-step adjustments, such as rate double robustness. We then study more complicated functionals such as dynamic treatment regimes, the linear-programming formulation for policy optimization in infinite-horizon Markov decision processes, and sensitivity analysis in causal inference. More broadly, we study optimization-based estimators, since this begets a class of estimands where identification via regression adjustment is straightforward but obtaining influence functions under minor variations thereof is not. The ability to approximate bias adjustments in the presence of arbitrary constraints illustrates the usefulness of constructive approaches for Gateaux derivatives.

Instrumental Variables

Source Condition Double Robust Inference on Functionals of Inverse Problems Vasilis Syrgkanis* Vasilis Syrgkanis, Andrew Bennett, Nathan Kallus, Xiaojie Mao, Whitney Newey, Masatoshi Uehara,

We consider estimation of parameters defined as linear functionals of solutions to linear inverse prob- lems. Any such parameter admits a doubly robust representation that depends on the solution to a dual linear inverse problem, where the dual solution can be thought as a generalization of the inverse propensity function. We provide the first source condition double robust inference method that ensures asymptotic normality around the parameter of interest as long as either the primal or the dual inverse problem is sufficiently well-posed, without knowledge of which inverse problem is the more well-posed one. Our result is enabled by novel guarantees for iterated Tikhonov regularized adversarial estima- tors for linear inverse problems, over general hypothesis spaces, which are developments of independent interest.

Machine Learning and Causal Inference

On the possibility of doubly robust root-n inference Matteo Bonvini* Matteo Bonvini, Edward H. Kennedy, Oliver Dukes, Sivaraman Balakrishnan,

We study the problem of constructing an estimator of the average treatment effect (ATE) that exhibits doubly-robust asymptotic linearity (DR-AL). This is a stronger requirement than doubly-robust consistency. In fact, a DR-AL estimator can yield asymptotically valid Wald-type confidence intervals even in the case when the propensity score or the outcome model is inconsistently estimated. On the contrary, the celebrated doubly-robust, augmented-IPW estimator requires consistent estimation of both nuisance functions for root-n inference. Previous authors have considered this problem (van der Laan, 2014, Benkeser et al, 2017, Dukes et al 2021) and provided sufficient conditions under which the proposed estimators are DR-AL. Such conditions are typically stated in terms of “high-level nuisance error rates” needed for root-n inference. In this paper, we build upon their work and establish sufficient and more explicit smoothness conditions under which a DR-AL estimator can be constructed. We also consider the case of slower-than-root-n convergence rates and clarify the connection between DR-AL estimators and those based on higher-order influence functions (Robins et al, 2017). We complement our theoretical findings with simulations.

Randomized Studies

Fusing efficiency: A review of data fusion methods with application to PIONEER 6 case study Xi Lin* Xi Lin, Jens Magelund Tarp, Robin Evans,

Integrating real-world data (RWD) and randomized controlled trials (RCTs) is becoming increasingly important in advancing causal inference in clinical research. This fusion holds great promise for enhancing the efficiency of average treatment effect estimation, thereby reducing the required number of trial participants and expediting drug access for patients in need. The FDA and EMA have recognized the complementary nature of these data sources and their integration to improve the quality of evidence. Despite the multitude of available data fusion methods, choosing the most suitable one for a specific research question is challenging. This difficulty arises from the diverse assumptions, associated limitations, and implementation complexities.
Our project aims to systematically review and compare data fusion methods, focusing on efficiency gain in average treatment effect (ATE) estimation. Through extensive simulations mirroring real-world scenarios, we identified a qualitative behaviour demonstrating a common risk-reward tradeoff across different methods. We investigate and interpret this tradeoff in various scenarios, providing important insights into understanding the strengths and weaknesses of different methods.
This presentation offers a comprehensive overview of available methods, highlights key findings from simulation studies and presents a real-world case study where PIONEER 6 trial is augmented with a US medical claims database for a more powerful ATE estimation.

Generalizability/Transportability

Multi-Source Conformal Inference Under Distribution Shift Larry Han* Larry Han, Yi Liu, Alexander Levis,

Conformal inference is a set of methods used to construct distribution-free, nonparametric prediction intervals with finite-sample marginal coverage guarantees. These methods have generally focused on covariate shift while assuming that conditional outcome distributions are invariant across environments. However, conditional outcome invariance is often violated in the real world. In this paper, we consider the problem of obtaining distribution-free prediction intervals for a target population leveraging multiple potentially biased data sources. Our approach is based on the efficient influence functions for the quantiles of unobserved outcomes in the target and source populations, combined with machine learning prediction algorithms to estimate nuisance functions, and a data-adaptive strategy to upweight informative data sources for efficiency gain and downweight non-informative data sources for bias reduction. We highlight the robustness and efficiency of our proposals for a variety of conformal scores and data-generating mechanisms via extensive synthetic experiments. We showcase the benefits of our method for obtaining more informative prediction intervals for pediatric patients undergoing high-risk cardiac surgery using data from 100 congenital heart centers in the United States.

Generalizability/Transportability

Data Fusion for Prospective and Retrospective Studies Ellen Graham* Ellen Graham, Andrea Rotnitzky, Marco Carone,

Previous work on data fusion has primarily focused on estimating parameters by leveraging data sources that align with variation-independent factors of the target population likelihood. In contrast, in this work, we introduce a general framework for debiased machine learning on smooth parameters by fusing a pair of data sources that align with variation-dependent components of the likelihood. Specifically, we consider the problem of data fusion when the distribution of the outcome given covariates (but not the covariate distribution) can be learned from a prospective cohort study and the distribution of the covariates given the outcome (but not the outcome distribution) can instead be learned from a retrospective case-control study. Our procedure allows for the identification of estimands that cannot be identified from either a prospective or retrospective study alone. We demonstrate how the dependence between these conditional distributions restricts the joint model, allowing for a reduction in the semiparametric efficiency bound. We characterize when estimators that achieve these bounds exist and provide a means to construct them. Finally, we provide examples of our proposed procedure for estimands of practical importance such as the average treatment effect and the average treatment effect on the treated.

Generalizability/Transportability

Constructing Synthetic Treatment Groups without the Mean Exchangeability Assumption
Yuhang Zhang* Yuhang Zhang, Yue Liu, Zhihua Zhang,

The purpose of this work is to transport the information from multiple randomized controlled trials to the target population where we only have the control group data. Previous works rely critically on the mean exchangeability assumption. However, as pointed out by many current studies, the mean exchangeability assumption might be violated. Motivated by the synthetic control method, we construct a synthetic treatment group for the target population by a weighted mixture of treatment groups of source populations. We estimate the weights by minimizing the conditional maximum mean discrepancy between the weighted control groups of source populations and the target population. We establish the asymptotic normality of the synthetic treatment group estimator based on the sieve semiparametric theory. Our method can serve as a novel complementary approach when the mean exchangeability assumption is violated. Experiments are conducted on synthetic and real-world datasets to demonstrate the effectiveness of our methods.

Heterogeneous Treatment Effects

Bayesian Causal Forests Combining Randomised And Observational Data For Heterogeneous Treatment Effects Estimation Ilina Yozova* Ilina Yozova, Ioanna Manolopoulou,

Bayesian Causal Forests (BCFs) are designed to estimate heterogeneous treatment effects using observational data by teasing apart the model into 3 pieces: prognostic effect – the influence of the covariates; treatment effect – the influence of the treatment; and propensity score, which captures the treatment assignment mechanism. However, when using data from different sources, the treatment assignment mechanism might differ greatly between each dataset. Additionally, the prognostic and/or treatment effects, as well as the sets of covariates, may also differ. A well-known example arises when combining randomised control trials (RCTs) and observational studies which can improve many aspects of causal inference, from increased statistical power to better external validity. Therefore, we extend the BCF model by introducing additional terms in the prognostic and treatment effects, which can absorb differences between two data sources, as well as capture some potential unobserved confounding of the observational data. Additionally, our model introduces a weighting parameter, allowing for adjustment of the influence of the observational data by raising to a power its contribution to the posterior distribution following a semi-modular inference approach. The flexibility of the power is valuable because it allows us to prevent RCT data from being swamped by the larger possibly confounded observational dataset. We implement our methods on a number of simulated and real data examples.

Sensitivity Analysis

A Calibrated Sensitivity Analysis for Weighted Disparity Decompositions Andy Shen* Andy Shen, Samuel Pimentel,

Disparities in health or well-being experienced by racial and sexual minority groups can be difficult to study using the traditional exposure-outcome paradigm in causal inference, since potential outcomes in variables such as race or sexual minority status are challenging to interpret. Decomposition analysis addresses this gap by considering causal impacts on a disparity under interventions to other, intervenable exposures (e.g. socioeconomic factors) that may play a mediating role in the disparity. While invoking weaker assumptions than causal mediation approaches, decomposition analyses are often conducted in observational settings and require uncheckable assumptions that rule out unmeasured confounders. Leveraging weighting estimators for disparity decomposition, we develop a sensitivity analysis for unobserved confounders in studies of disparities using the marginal sensitivity model. We use the percentile bootstrap to construct valid confidence intervals for disparities and causal effects on disparities under given levels of confounding under mild conditions. We also explore amplifications that give insight into multiple confounding mechanisms. We illustrate our framework on a study examining disparities in youth suicide rates among sexual minorities using the Adolescent Brain Cognitive Development Study (ABCD). Supported by the National Science Foundation under Grant No. 2142146.

Sensitivity Analysis

Causal progress with imperfect placebo treatments and outcomes Chad Hazlett* Chad Hazlett, Adam Rohde,

In the quest to make defensible causal claims from observational data, investigators may leverage information from “placebo treatments” and “placebo outcomes” (aka “negative control outcomes”). Existing approaches focus largely on point identification and require two difficult assumptions: (i) “perfect placebos” (placebo treatments have precisely zero effect on the outcome; treatment has precisely no effect on placebo outcomes); and (ii) “equiconfounding” (the treatment-outcome relationship where one is a placebo suffers the same amount of confounding as does the real treatment-outcome relationship, on some scale). By contrast, using an omitted variable bias framework, we consider degrees of placebo imperfection (non-zero effects of placebo treatment on real outcomes or of real treatments on placebo outcomes), and non-equiconfounding (different strengths of confounding suffered by a placebo treatment/outcome compared to the true treatment-outcome relationship). Postulated values for these quantities identify or bound the linear estimates of treatment effects. While applicable in many settings, one broad use-case for this approach is to employ pre-treatment outcomes as (perfect) placebo outcomes. In this setting, our approach offers a credibility-enhancing relaxation of the parallel trends assumption of DID. We demonstrate the use of our framework with two applications, employing an R package that implements these approaches.

Sensitivity Analysis

Automatically Calibrated Sensitivity Models for Causal Inference with Unmeasured Confounding Alexander McClean* Alexander McClean, Zach Branson, Edward Kennedy,

Accounting for unmeasured confounding is crucial when estimating causal effects with observational data. For this purpose, we propose several data-driven methods based on novel automatically calibrated sensitivity (ACS) models, which bound the error due to unmeasured confounding by an analogous notion of error due to measured confounding, multiplied by a sensitivity parameter. We illustrate how to construct ACS models via several examples and demonstrate their advantages over standard sensitivity and post hoc calibration analyses. We focus on estimating a one-number summary of Average Treatment Effect sensitivity — an intuitive alternative to more frequently considered estimands in the literature — with ACS models defined at the level of (1) the causal effect, (2) the counterfactual outcome regressions, and (3) the odds ratio of the probability of receiving treatment. Under all models, we observe that either a margin condition or smooth approximation is required for efficient estimation and develop methods for estimation and inference with both. Moreover, we establish that our estimators are doubly robust, and attain parametric efficiency and asymptotic normality under nonparametric conditions on the relevant nuisance function estimators. Finally, we illustrate our methods with two data analyses, examining the effect of exposure to violence on attitudes towards peace and mothers' smoking on infant birth weight.

Generalizability/Transportability

Sensitivity Analysis for Extending Inferences of a Binary Time-Fixed Treatment Effect Derived from a Randomized Controlled Trial to a Target Population When a Subset of Treatment Effect Modifiers Are Measured Only On Trial Subjects Jay Xu* Jay Xu, Marissa Seamans,

The external validity of results from RCTs may be compromised when the distribution of effect modifiers (EMs) differs between the study population and the target population of scientific and/or policy interest. To better estimate real world and/or policy relevant causal effects of treatments or interventions, integrative methods that synthesize data from RCTs and observational data sources (e.g., EHR data) to infer causal effects for target populations represented by these observational data sources have been developed, which explicitly adjust for observed differences in covariate distributions between RCT and observational samples. Their utility of such methods, however, depends on the extent to which EMs that differ in distribution between the study and target populations are measured in both RCT and observational samples. We consider the theoretical scenario where all EMs are measured in the RCT sample, but a subset of them are unmeasured in the observational sample. We propose sensitivity analysis procedures to perform Frequentist or Bayesian inference for the target population average treatment effect of a binary time-fixed treatment under various user-postulated differences between the study and target population conditional distributions of the EMs measured only in the RCT given the dually measured covariates. We demonstrate the merits of the proposed sensitivity analysis procedures using a simulation study and illustrate their use on substance use clinical trial data.

Sensitivity Analysis

Sensitivity Analysis to Unobserved and Residual Confounding in the Effect of Physical Activity on Mortality among Former Smokers Rui Hu* Rui Hu, Charles Matthews, Neal Freedman, Maki Inoue-Choi, John Staudenmayer, Ted Westling,

Recent research suggests that physical activity is associated with reduced risk of mortality due to respiratory disease and cancer among former smokers after adjusting for common causes using data from the NIH-AARP Study. This study measured former smoking behavior using self-reported average number of cigarettes smoked per day (CPD), which may have measurement error and may not fully reflect previous smoking behavior as length of time spent smoking was not recorded. As previous smoking behavior causes respiratory disease and lung cancer, these associations may be biased estimates of the true causal effects in either of these cases. Determining whether these effects are causal is important, since former smokers want to know if they can reduce their risk of these diseases by exercising more. We compare two types of causal sensitivity analyses: to measurement error in CPD, and to unobserved confounding. We find that the effect of physical activity on respiratory disease mortality is not explained away by a moderate amount of unobserved confounding or high measurement error. The effect of physical activity on lung cancer is explained away by a small amount of unobserved confounding, but not by measurement error. We hypothesize that the robustness to measurement error could be due to assumptions of the measurement error model, and we discuss the implications of these results for using standard measurement error models in causal sensitivity analyses.

Machine Learning and Causal Inference

Non-parametric efficient estimation of marginal structural models with multi-valued time-varying treatments Axel Martin* Axel Martin, Ivan Diaz, Michele Santacatterina,

Marginal structural models are a popular method for estimating causal effects in the presence of time-varying exposures. In spite of their popularity, no scalable non-parametric estimator exist for marginal structural models with multi-valued and time-varying treatments. In this paper, we use machine learning together with recent developments in semiparametric efficiency theory for longitudinal studies to propose such an estimator. The proposed estimator is based on a study of the non-parametric identifying functional, including first order von-Mises expansions as well as the efficient influence function and the efficiency bound. We show conditions under which the proposed estimator is efficient, asymptotically normal, and sequentially doubly robust in the sense that it is consistent if, for each time point, either the outcome or the treatment mechanism is consistently estimated. We perform a simulation study to illustrate the properties of the estimators, and present the results of our motivating study on semi-synthetic and trial data.

Machine Learning and Causal Inference

Geometry-Aware Normalizing Wasserstein Flows for Optimal Causal Inference Kaiwen Hou* Kaiwen Hou,

Introduction:
We introduce a novel approach for causal inference that integrates continuous normalizing flows (CNFs) with Wasserstein gradient flows. This method improves upon traditional TMLE by incorporating geometric awareness in model navigation, focusing on minimizing the Cramér-Rao bound from p_0 and p_1. Our approach combines the versatility of CNFs with the rich geometric structure of the 2-Wasserstein metric, enhancing both the flexibility and accuracy of causal effect estimates.
Theory and Methods:
The heart of our approach is the transformation of simple base distributions into complex targets using CNFs. We propose normalizing Wasserstein flows that optimize CNF parameters for minimal discrepancy between modeled and target distributions, ensuring invertibility and smoothness. Key to our framework is variance regularization and a generalized formulation focusing on velocity field alignment in the loss function, simplifying computational demands.
Optimal Causal Inference:
Addressing distribution shifts and biases in estimating population parameters, our method is applied to optimal causal inference, with an emphasis on local semiparametric efficiency. By minimizing the efficiency bound across the model manifold trajectory, our approach aims at efficient estimators in finite-sample scenarios. Preliminary experimental results show reduced mean-squared error and variance of efficient influence functions compared to traditional methods like TMLE and AIPW.

Machine Learning and Causal Inference

Interpretable Causal Inference for Analyzing Wearable, Sensor, and Distributional Data
Srikar Katta* Srikar Katta, Harsh Parikh, Alexander Volfovsky, Cynthia Rudin,

Many modern causal questions ask how treatments affect complex outcomes that are measured using wearable devices and sensors. Current analysis approaches require summarizing these data into scalar statistics (e.g., the mean), but these summaries can be misleading. For example, disparate distributions can have the same means, variances, and other statistics. Researchers can overcome the loss of information by instead representing the data as distributions. We develop an interpretable method for distributional data analysis that ensures trustworthy and robust decision-making: Analyzing Distributional Data via Matching After Learning to Stretch (ADD MALTS). We (i) provide analytical guarantees of the correctness of our estimation strategy, (ii) demonstrate via simulation that ADD MALTS outperforms other distributional data analysis methods at estimating treatment effects, and (iii) illustrate ADD MALTS' ability to verify whether there is enough cohesion between treatment and control units within subpopulations to trustworthily estimate treatment effects. We demonstrate ADD MALTS' utility by studying the effectiveness of continuous glucose monitors in mitigating diabetes risks.

Machine Learning and Causal Inference

Assumption-Lean Quantile Regression Georgi Baklicharov* Georgi Baklicharov, Christophe Ley, Stijn Vansteelandt,

Quantile regression is a powerful tool for detecting exposure-outcome associations given covariates across different parts of the outcome’s distribution, but has two major limitations when the aim is to infer the effect of an exposure. Firstly, the exposure coefficient estimator may not converge to a meaningful quantity when the model is misspecified, and secondly, variable selection methods may induce bias and excess uncertainty, rendering inferences biased and overly optimistic. In this paper, we address these issues via partially linear quantile regression models which parametrize the conditional association of interest, but do not restrict the association with other covariates in the model. We propose consistent estimators for the unknown model parameter by mapping it onto a nonparametric main effect estimand that captures the (conditional) association of interest, even when the quantile model is misspecified. This estimand is estimated using the efficient influence function under the nonparametric model, allowing for the incorporation of data-adaptive procedures such as variable selection and machine learning. Our approach provides a flexible and reliable method for detecting associations that is robust to model misspecification and excess uncertainty induced by variable selection methods.

Machine Learning and Causal Inference

Debiasing Machine-Learning- or AI-Generated Regressors in Partial Linear Models Jingwen Zhang* Jingwen Zhang, Wendao Xue, Yifan Yu, Yong Tan,

Researchers are increasingly leveraging machine learning (ML) or artificial intelligence technologies (AI) to predict feature variables and use them as regressors in subsequent econometric models. However, because ML/AI predictions are imperfect, these generated regressors would inevitably contain measurement errors. The direct use of such regressors in subsequent econometric models can result in biased estimation, ultimately leading to inaccurate conclusions. In light of this, we examine the problem of debiasing ML/AI-generated regressors in both linear and partial linear regression models. We propose estimators that utilize the Two-Stage Least Square (TSLS) and the Generalized Method of Moments (GMM) under the Double Machine Learning (DML) framework. We demonstrate the asymptotic consistency and normality of our estimators and conduct extensive Monte Carlo simulations to show the outperformance of our estimators compared with other methods. Our work advances causal inference in addressing measurement error problems arising from ML/AI-generated regressors in partial linear models. Our work provides valuable practical implications for designing experimental systems and overcoming ML/AI biasedness.

Machine Learning and Causal Inference

Inverting Estimating Equations for Causal Inference on Quantiles Chao Cheng* Chao Cheng, Fan Li,

The causal inference literature frequently focuses on estimating the mean of the potential outcome, whereas the quantiles of the potential outcome may carry important additional information. We propose a universal approach, based on the inverse estimating equations, to generalize a wide class of causal inference solutions from estimating the mean of the potential outcome to its quantiles. We assume that an identifying moment function is available to identify the mean of the threshold-transformed potential outcome, based on which a convenient construction of the estimating equation of quantiles of potential outcome is proposed. In addition, we also give a general construction of the efficient influence functions of the mean and quantiles of potential outcomes, and identify their connection. We motivate estimators for the quantile estimands with the efficient influence function, and develop their asymptotic properties when either parametric models or data-adaptive machine learners are used to estimate the nuisance functions. A broad implication of our results is that one can rework the existing result for mean causal estimands to facilitate causal inference on quantiles, rather than starting from scratch. Our results are illustrated by several examples, including assessing quantile treatment effect with a single or time-varying treatment, quantile mediation analysis, and principal stratification.

Heterogeneous Treatment Effects

Synthetic Combinations: A Causal Inference Framework for Combinatorial Interventions
Abhineet Agarwal* Abhineet Agarwal, Anish Agarwal, Suhas Vijaykumar,

We consider a setting where there are N heterogeneous units and p interventions. Our goal is to learn unit-specific potential outcomes for any combination of these p interventions, i.e., N x 2^p causal parameters. Choosing a combination of interventions is a problem that naturally arises in a variety of applications such as factorial design experiments (e.g., multivariate tests on digital platforms), combination therapies in medicine, rankings in recommendation engines, etc. Running N x 2^p experiments to estimate all parameters is likely infeasible as N and p grow. Further, with observational data there is likely confounding. To address these challenges, we propose a novel latent factor model that imposes structure across units (i.e., the matrix of potential outcomes is rank r) and combinations of interventions (i.e., Fourier expansion of the potential outcomes is s sparse). We establish identification for all N x 2^p parameters despite unobserved confounding. We propose an estimation procedure, Synthetic Combinations, and establish it is finite-sample consistent and asymptotically normal. Synthetic Combinations is consistent given poly(r) x (N + s^2p) observations, while previous methods have sample scaling as min(N x s^2p, poly(r) x (N + 2^p)). We also use Synthetic Combinations to propose a data-efficient experimental design mechanism. Empirically, we show Synthetic Combinations outperforms competing approaches on a real-world dataset on movie recommendations.

Heterogeneous Treatment Effects

Causal inference with constraints on the probability of intervention Alexander Levis*
Alexander Levis, Eli Ben-Michael, Edward Kennedy,

Treatment rules or policies are mappings from individual patient characteristics to tailored treatment assignments. Optimal policies that maximize mean outcomes have been well characterized in unconstrained settings, in cases where one treatment level is limited in supply, and under cost constraints when treatment cost is random. In this work, we describe a novel resource-limited setting, important for applications in health policy, in which treatment options are freely accessible but the ability to intervene on a portion of a target population is constrained, e.g., if the population is large, and follow-up and encouragement of treatment uptake is labor-intensive. We derive formulas for optimal treatment rules in such settings, and for any given budget, quantify the loss compared to the optimal unconstrained rule. We then propose efficient and robust influence function-based estimators of the mean outcome under the optimal constrained rule, and other related quantities that are of independent interest beyond this resource-limited setting. Finally, we demonstrate our framework in simulations and in a longitudinal observational study.

Heterogeneous Treatment Effects

Modeling Time-Varying Effects of Mobile Health Interventions Using Longitudinal Functional Data from HeartSteps Micro-Randomized Trial Jiaxin Yu* Jiaxin Yu, Predrag Klasnja, Susan Murphy, Tianchen Qian,

Understanding how the effect of a mobile health intervention varies over time and with contextual information is critical for both optimizing the intervention and advancing domain knowledge. This analysis aims to assess how a push notification suggesting physical activity influences individuals' step count and how such influence varies over time, using data from the HeartSteps micro-randomized trial (MRT). The statistical challenges include the time-varying treatments and the longitudinal functional step count measurements. We propose the first semiparametric causal excursion effect model with varying coefficients to model the time-varying effects within a decision point and across decision points in an MRT. The proposed model incorporates double time indices to accommodate the longitudinal functional outcome, enabling the assessment of time-varying effect moderation by contextual variables. We propose a two-stage causal effect estimator that is robust against a misspecified high-dimensional outcome regression model. We establish asymptotic theory and conduct simulation studies to validate the proposed estimator. Our analysis provides new insights into individuals' change in response profiles (such as how soon a response occurs) due to the activity suggestions, how such changes differ by the type of suggestion they receive, and how such changes depend on other contextual information such as being recently sedentary and the day being a weekday.

Weighting

A Robust Sequential Covariate Balancing Approach for Estimating the Effects of Time-Varying Treatments on Survival Outcomes Yige Li* Yige Li, José Zubizarreta,

In longitudinal studies, treatments or exposures can vary across time and depend on covariates responding to previous exposures. For such studies, we propose Robust Sequential Covariate Balancing (RSCB), a flexible and stable weighting technique, designed to estimate the effects of time-varying treatments on a general class of outcomes, including survival outcomes. RSCB utilizes a backward covariate balancing procedure for identification and estimation. RSCB has a product form parallel to inverse probability weighting (IPW) but utilizes modellable trends in the covariates and meanwhile minimizes weights variation. Unlike IPW methods, whose estimates can converge at slow rates because of inadequate covariate overlap, RSCB gains efficiency by prioritizing outcome-relevant forms of covariate balance. In comparison to the g-computation formula, RSCB does not extrapolate and maintain robustness to misspecification of covariates and outcomes models. In contrast with longitudinal stable balance weighting (LSBW), RSCB can accommodate multiple types of outcomes, longer time courses, and higher covariate dimensions. We illustrate this new method in a study of the effects of peer antisocial behaviors on drug relapse of adolescents.

Weighting

Targeted Function Balancing Leonard Wainstein* Leonard Wainstein,

This paper introduces Targeted Function Balancing (TFB), a covariate balancing weights framework for estimating the average treatment effect of a binary intervention. TFB first regresses an outcome on covariates, and then selects weights that balance functions (of the covariates) that are probabilistically near the resulting regression function. This yields balance in the regression function's predicted values and the covariates, with the regression function's estimated variance determining how much balance in the covariates is sufficient. Notably, TFB demonstrates that intentionally leaving imbalance in some covariates can increase efficiency without introducing bias, challenging traditions that warn against imbalance in any variable. Additionally, TFB is entirely defined by a regression function and its estimated variance, turning the problem of how best to balance the covariates into how best to model the outcome. Kernel regularized least squares and the LASSO are considered as regression estimators. With the former, TFB contributes to the literature of kernel-based weights. As for the LASSO, TFB uses the regression function's estimated variance to prioritize balance in certain dimensions of the covariates, a feature that can be greatly exploited by choosing a sparse regression estimator. This paper also introduces a balance diagnostic, Targeted Function Imbalance, that may have useful applications.

Weighting

Estimating Causal Effects for Binary Outcomes Using Per-Decision Inverse Probability Weighting Yihan Bao* Yihan Bao, Tianchen Qian,

Micro-randomized trials are commonly conducted for optimizing mobile health interventions such as push notifications for behavior change. In analyzing such trials, causal excursion effects are often of primary interest, and their estimation typically involves inverse probability weighting (IPW). However, in a micro-randomized trial, additional treatments can often occur during the time window over which an outcome is defined, and this can greatly inflate the variance of the causal effect estimator because IPW would involve a product of numerous weights. To reduce variance and improve estimation efficiency, we propose a new estimator using a modified version of IPW, which we call “per-decision IPW”. It is applicable when the outcome is binary and can be expressed as the maximum of a series of sub-outcomes defined over sub-intervals of time. We establish the estimator’s consistency and asymptotic normality. Through simulation studies and real data applications, we demonstrate substantial efficiency improvement of the proposed estimator over existing estimators (relative efficiency up to 1.45 and sample size savings up to 31% in realistic settings). The new estimator can be used to improve the precision of primary and secondary analyses for micro-randomized trials with binary outcomes. We applied the new estimator to different MRT datasets (HeartSteps, Drink Less) to analyze the moderator variables for the time-varying causal effects of different interventions.

Weighting

Towards representation learning for general weighting problems in causal inference Oscar Clivio* Oscar Clivio, Avi Feller, Chris Holmes,

Weighting problems in treatment effect estimation can be solved by minimising an appropriate probability distance. Choosing which distance to minimise, however, can be challenging as it depends on the unknown data generating process. An alternative is to instead choose a distance that depends on a suitable representation of covariates. In this work, we give errors that quantify the bias added to a weighting estimator when using a representation, giving clear objectives to minimise when learning the representation and generalising a large body of previous work on deconfounding, prognostic, balancing and propensity scores. We further outline a method minimising such objectives, and show promising numerical results on two semi-synthetic datasets.

Weighting

The Role of the Simplex Constraint in Regularizing Treatment Effect Estimates David Arbour* Avi Feller, David Arbour, Anup Rao, Tung Mai,

Many workhorse methods in causal inference are weighting estimators, with estimates that are linear in the observed outcomes. A key consideration for these methods is whether to constrain the weights to be non-negative (or on the simplex): matching and inverse propensity score weighting (IPW), for example, impose this constraint while linear regression does not. These constraints limit extrapolation but can introduce bias, especially with high-dimensional features. In this paper, we take a geometric perspective and argue that the simplex constraint acts as an implicit regularizer via sample trimming. We make this regularization explicit as a form of generalized ridge penalty, characterize the resulting behavior under a high-dimensional factor model, and consider the mis-specified setting, such as when the target is outside the convex hull of control units. We then extend results to other shape constraints on the weights, especially to popular forms of Augmented IPW. We argue that, beyond their recognized statistical properties, these estimators also have attractive geometric properties, especially in high dimensions. Finally, we show how this perspective helps explain recent results on high-dimensional causal inference, such as the lack of double-descent behavior for the synthetic control method.

Mediation

A Covariance Perspective on Randomized Interventional Analogues of Mediation Estimands
Ang Yu* Ang Yu, Li Ge,

In causal mediation analysis, the natural indirect effect (NIE) typically requires the cross-world independence assumption for identification, an assumption often unrealistic in many settings. Alternatively, the randomized interventional analogue (RIA) of the NIE circumvents this assumption. However, Miles (2023) demonstrates through specific counterexamples that the RIA of the NIE fails to meet certain null criteria essential for a valid indirect effect measure. This paper elucidates that the discrepancy between the NIE and its RIA is representable as a covariance between the mediator and outcome. Specifically, when both are binary, this difference equates to the covariance between the treatment's effect on the mediator and the mediator's effect on the outcome. Thus, we demystify the violation of the null criteria, detail the conditions under which this occurs, and highlight the uniquity of such violation. Similarly, we examine the differences between the natural direct effect, the total effect, and their RIAs. Furthermore, our work also enriches the extensive literature on causal U-statistics, including Wilcoxon-Mann-Whitney parameters, win ratio, and probability of causation. We observe that the estimands commonly used in practice are frequently misinterpreted and mistaken for their more intuitively interpretable counterparts. We establish that the differences between these commonly used estimands and their misinterpretations can also be represented as covariances.

Mediation

Unpacking subgroup differences in treatment effects: A causal decomposition approach for mediated moderation analysis Xiao Liu* Xiao Liu,

Assessing differences between demographic subgroups (e.g., female and male) in treatment effect—or, moderation analysis—is important in behavioral sciences. In moderation analysis, besides quantifying how much subgroups differ in treatment effect ("total moderation"), it is often useful to examine why the effect differences between subgroups arise—such as by examining intermediate variables contributing to the effect difference—or, mediated moderation analysis.

For causal inference involving intermediates, causal mediation methods are fast-growing but have limited development for mediated moderation analysis; a particular challenge is that the subgroups are often defined by demographic characteristics, which are non-manipulable.

This study extends the causal decomposition approach, and develops methods for mediated moderation analyses with causal interpretation. Our methods decompose the total moderation into the causal estimands capturing how much the subgroup difference in the effect is attributable to the subgroup difference in intermediate variable(s) ("mediated moderation") and how much is not ("remaining moderation"). We develop multiply-robust estimators (including cross-fitted one-step estimators and targeted minimum loss estimators), which facilitate using machine learning techniques in causal inference. We illustrate the applications in an empirical mediated moderation analysis to unpack gender differences in the effects of an intervention for behavior problems.

Mediation

Identification and estimation of mediational effects of longitudinal modified treatment policies Brian Gilbert* Brian Gilbert, Katherine Hoffman, Nicholas Williams, Kara Rudolph, Iván Díaz, Edward Schenck,

We put forward a comprehensive semiparametric approach to causal mediation analysis, addressing the complexities inherent in settings with longitudinal and continuous treatments. Our methodology utilizes a nonparametric structural equation model and a sequential regression technique, yielding an efficient estimator without relying on restrictive modeling assumptions. We are motivated by a recent scientific controversy regarding the effects of invasive mechanical ventilation (IMV) on the survival of COVID-19 patients in the ICU, considering acute kidney injury (AKI) as a mediating factor. We highlight the possibility of “inconsistent mediation,” in which the direct and indirect effects of the exposure operate in opposite directions. We discuss the significance of mediation analysis for scientific understanding and its potential utility in treatment decisions.

Algorithmic Causal Inference

An Efficient Algorithm for Closed-Form Partial Identification of Causal Quantities
Guilherme Duarte* Guilherme Duarte,

Since the groundbreaking contributions of Manski (1990) and Balke and Pearl (1994) concerning bounds on causal quantities, scholars have directed their attention toward scenarios marked by the non-existence of point-identification solutions. Rather than capitulating or modifying their original inquiries to align with more tractable estimands, researchers have endeavored to derive ranges of potential values that conform to their assumptions and empirical data. For instance, in experimental settings characterized by imperfect compliance, the reliance on estimands like the Local Average Treatment Effect – which is point-identifiable but lacks a straightforward interpretation – can be supplanted by the estimation of sharp bounds for the non-identifiable Average Treatment Effect. However, while partial identification represents a preferable strategy, its implementation often proves challenging. Complete algorithmic solutions typically resort to numeric methods (Duarte et al., 2023), introducing complications for subsequent inference. To address these challenges, we present an efficient algorithm capable of providing closed-form solutions for causal-complete problems with symmetric constraints. These problems encapsulate a substantial portion of problems encountered in causal inference. Significantly, our algorithm not only cuts down on computation time compared to existing solutions but also unveils novel solutions previously unknown.

Bayesian Causal Inference

A Bayesian Classification Trees Approach to Treatment Effect Variation with Noncompliance Jared Fisher* Jared Fisher, David Puelz, Sameer Deshpande,

Estimating varying treatment effects in randomized trials with noncompliance is inherently challenging since variation comes from two separate sources: variation in the impact itself and variation in the compliance rate. In this setting, existing frequentist and machine learning methods are quite flexible but are highly sensitive to the so-called weak instruments problem, in which the compliance rate is (locally) close to zero. Parametric Bayesian approaches, which account for noncompliance via imputation, are more robust in this case, but are much more sensitive to model specification. In this paper, we propose a Bayesian machine learning approach that combines the best features of both approaches. Our main methodological contribution is to present a Bayesian Causal Forest model for binary response variables in scenarios with noncompliance. In this Bayesian noncompliance framework, we repeatedly impute individuals' compliance types, allowing us to flexibly estimate varying treatment effects among compliers. Binary outcomes add a layer of complexity as estimation involves the probability of occurrence, not the binary outcome, and these probabilities are not observed. We apply the method to detect and analyze heterogeneity in a study of workplace wellness, where there are a plethora of binary outcomes of interest.

Bayesian Causal Inference

Density Regression Bayesian Causal Forests Anna Morgan* Anna Morgan, Jared Murray,

We introduce a nonparametric Bayesian approach for estimating heterogeneous effects with density regression, building upon Bayesian Causal Forests (BCF) and prior Bayesian Additive Regression Trees (BART) density regression methods. We incorporate a targeted smoothing prior on terminal tree nodes to ensure smoothness and a reduced-rank kernel approximation to make (approximate) Gaussian process regression computationally feasible. Allowing distributional features of the response density to vary with covariates relaxes strong assumptions about the treatment effect mechanism and allows for richer insights about treatment effect heterogeneity. We illustrate our model by applying it to data from a recent high-profile mindset intervention experiment, allowing new exploration of differential treatment effect patterns and subgroups of interest. Finally, we carefully consider prior specification and the implied regularization on treatment effects of interest to tune our model for causal inference in a realistic setting. We propose reasonable default priors for parameters of the treatment effect function and discuss how these priors may be informed by beliefs about the potential scale of treatment effects, illustrated by simulation studies.

Bayesian Causal Inference

Synthetic control method for clinical trials and precision medicine Veronica Ballerini*
Veronica Ballerini, Giulio Grossi,

The large data availability fuels research on “one patient, one treatment” personalized medicine methods in clinical studies, particularly impactful for rare diseases where randomized controlled trials are impractical. Existing approaches like “external controls” using data from similar patients can be problematic in cases where there are limited suitable controls. This work proposes a novel approach combining historical controls with an adaptive study design, allowing treatment assignment to adapt based on valid estimates of treatment effects obtained within patient subpopulations. The proposed method represents a novel use of the “econometric” synthetic control method within clinical studies. We use a Bayesian approach to inference, which makes it easy to quantify the uncertainty related to the estimates and propagate it to the next patient’s treatment assignment. The novel combination of synthetic controls with adaptive designs leverages the flexibility of both methods for the effective handling of complex data patterns, improving treatment allocation efficiency while guaranteeing ethical studies. Simulation results support the promise of this approach for clinical studies.

Causal Discovery

Towards Causal Discovery with Statistical Guarantees Shreya Prakash* Shreya Prakash, Fan Xia, Elena Erosheva,

Causal discovery methods aim to determine the causal direction between variables using data, such as whether sleep problems cause depression or vice versa. Causal discovery algorithms, like LiNGAM, use structural and distributional assumptions to estimate the causal direction. However, these algorithms often lack a way to measure uncertainty in their estimates or their finite-sample performance when assumptions are violated. We introduce the True Direction Detection Rate (TDDR) metric and a TDDR-based procedure to quantify uncertainty and assess the finite sample performance of causal discovery methods. The TDDR calculates the probability of accurately predicting the true causal direction as a function of sample size. We demonstrate the TDDR in a bivariate linear causal discovery setting and show its applicability to various causal discovery methods. These include methods based on independence measure comparisons like LiNGAM and those based on hypothesis testing like our proposed test-based method, which provides more statistical guarantees compared to the former. Through simulations, we validate asymptotic normality for the TDDR and demonstrate its use for causal discovery methods like LiNGAM and the test-based method when linearity and non-Gaussianity assumptions are violated. Our work provides insights into assessing the finite-sample performance and uncertainty of causal discovery methods, especially when assumptions are not met.

Causal Discovery

Causal Discovery in Directed, Possibly Cyclic, Graphical Models Pardis Semnani* Pardis Semnani, Elina Robeva,

We consider the problem of learning a directed graph G* from observational data. We assume that the distribution which gives rise to the samples is Markov and faithful to the graph G* and that there are no unobserved variables. We do not rely on any further assumptions regarding the graph or the distribution of the variables. In particular, we allow for directed cycles in G* and work in the fully non-parametric setting. Given the set of conditional independence statements satisfied by the distribution, we aim to find a directed graph which satisfies the same d-separation statements as G*. We propose a hybrid approach consisting of two steps. We first find a partially ordered partition of the vertices of G* by optimizing a certain score in a greedy fashion. We prove that any optimal partition uniquely characterizes the Markov equivalence class of G*. Given an optimal partition, we propose an algorithm for constructing a graph in the Markov equivalence class of G* whose strongly connected components correspond to the elements of the partition, and which are partially ordered according to the partial order of the partition. Our algorithm comes in two versions — one which is provably correct and another one which performs fast in practice.

Causal Discovery

Root cause discovery Jinzhou Li* Jinzhou Li, Benjamin Chu, Ines Scheller, Julien Gagneur, Marloes Maathuis,

We study the problem of root cause discovery, which aims to find the intervened variable of one interventional sample using observational samples as a reference. This problem is motivated by applications such as discovering the first genetic variant for rare disease patients. We consider a linear structural equation model where the causal ordering is unknown. We start by studying a simple method based on marginal squared z-scores, and we characterize when this method can identify the root cause and when it will fail. We then prove, without any additional assumptions, that the root cause is identifiable by the Cholesky decomposition, even when the causal ordering is unidentifiable. This identifiability result inspires a method for root cause discovery that requires p! permutations, where p is the number of variables. Hence, it is not feasible for applications with a large p. To address this issue, we characterize which permutations can result in the correct root cause. This result leads to a new method for root cause discovery requiring at most p permutations. At last, we examine the performance of our methods in simulations.

Causal Inference and Bias/Discrimination

Credible Evidence of Gender Discrimination Using Instrumental Inequality Jiwoo Kim* Jiwoo Kim, Yongnam Kim,

Considering sex discrimination as a direct effect of sex on an outcome given a justifiable mediator, one of the significant challenges for a statistical approach to measuring sex discrimination is the presence of mediator-outcome confounding. As in Bickel et al.'s (1979) example of sex bias in UC Berkeley, which compares males' and females' acceptance rates in each department, the conventional approach is susceptible to potential confounders that may influence both department choice (mediator) and acceptance (outcome). This paper introduces a novel approach using the causal constraint of instrumental inequality to investigate sex discrimination. This method has the advantage of requiring neither the absence of mediator-outcome confounding nor specific functional forms or distributional assumptions. Applying the proposed approach to two sex discrimination cases in South Korea, this paper offers more credible statistical evidence of sex discrimination without relying on unrealistic assumptions.

Causal Inference and Bias/Discrimination

Nonparametric Inference on Dose-Response Curves Without Positivity Condition Yikun Zhang* Yikun Zhang, Yen-Chi Chen, Alexander Giessing,

This paper presents a novel integral estimator for estimating the dose-response curve in the presence of spatial confounding without requiring the commonly assumed positivity condition. Our approach involves estimating the derivative of the treatment effect and integrating it to address the bias resulting from the lack of positivity condition. We also provide a fast and reliable numerical recipe for approximating our estimator and derive related asymptotic theory.
To account for uncertainty, we propose a bootstrap method that generates a simultaneous confidence band for the dose-response curve. Additionally, we propose an inverse probability weighting approach for estimating the derivative effect under the violation of positivity conditions, which exhibits connections to the support estimation problem.

Causal Inference and Bias/Discrimination

Data-Adaptive Experimentation to Find Contexts with the Most and Least Discrimination
Jennah Gosciak* Jennah Gosciak, Daniel Molitor, Ian Lundberg,

Randomized experiments reveal discriminatory choices. From audit studies to online experiments, designs follow a common recipe: present a decision-maker with a profile that holds constant the context (e.g., the political experience of a candidate) and randomizes a signal (e.g., the age of the candidate). This design isolates the causal effect of the signal on the choice: the decision must be caused by age. But what if discrimination differs across contexts? Voters might prefer a younger candidate only if that candidate has political experience, for example. Experiments that fix different contexts would detect different amounts of discrimination. Standard designs cannot discover this variation because the context is fixed. Conjoint experiments cannot discover this variation because they are powered for average marginal effects. By applying Thompson sampling methods, we data-adaptively discover contexts with high and low levels of discrimination. We illustrate with two new pre-registered online experiments. Our first experiment (completed) explores how the effect of candidate age on voter preferences depends on the context of candidate race, gender, and experience. Our second experiment (funded and IRB approved) explores how hiring discrimination against mothers may depend on the context of the applicant’s race and educational credentials. Our goal is to show how experiments can not only detect discrimination, but adaptively discover the contexts where it is most severe.

Causal Inference and Bias/Discrimination

Machine Learning Regression Adjustment for RCTs with Survey Outcomes Alex Whitworth* Alex Whitworth,

Randomized Controlled Trials (RCTs) with survey outcomes are common in industry. These studies present two challenges to the practitioner: (i) the survey completion population may be unbalanced compared to the treated population; and (ii) there may be systematic biases in survey non-completion. In this work, we implement and evaluate a generalized Oaxaca-Blinder estimator (Guo, Basse; 2023) using post-stratification weighting. We evaluate this estimator via an extension of the dowhy framework (Sharma, Kiciman; 2020) where we simulate binary outcomes with binary survey completion. The Oaxaca-Blinder estimator is compared to the standard difference in means estimator. We find that the Oaxaca-Blinder estimator produces unbiased estimates with ~1/3rd narrower confidence intervals and is robust to (i) correlation of survey completion with survey outcome; (ii) unobserved confounders; and (iii) data subset validation. However our simulations show some problems with placebo tests. This work is consistent with prior work showing that variance reduction in RCTs is achievable using machine learning regression adjustment estimators and that the variance reduction is achievable for RCTs with survey outcomes.

Causal Inference and Bias/Discrimination

Targeted Learning Estimation of Sampling Variance for Improved Inference Yunwen Ji*
Yunwen Ji, Mark van der Laan, Alan Hubbard,

Variance estimation is crucial for robust statistical inference in causal inference. Existing research highlights the potential finite-sample limitations of standard errors based on sample variance estimates of plug-in influence curves (IC) for asymptotically linear estimators. The IC-based variance estimator often yields anti-conservative estimates, leading to elevated Type-I error rates and poor coverage, particularly in limited samples or in the presence of near positivity violations. In this paper, we address this challenge through the introduction of a one-step targeted variance estimator of causal risk ratio (CRR) in settings involving treatment, outcome, and baseline covariates. The paper focuses primarily on the variance of log(CRR), but the findings can be easily generalized to other causal effect parameters. Specifically, the parameter of interest for our research is the variance of the IC for an estimator of log(CRR). Several methods of developing efficient estimators of asymptotically linear parameters are available. In this paper, we concentrate on the so-called one-step targeted maximum likelihood estimator, which is a substitution estimator that utilizes a one-dimensional universal least favorable parametric submodel when updating the distribution. Results from simulations under finite sample sizes and high positivity showcase the advantages of our variance estimator, exhibiting a more robust estimation with a coverage closer to 0.95 and lower Type-I errors.

Causal Inference and SUTVA/Consistencies Violations

Modeling Interference Using Experiment Roll-out Ariel Boyarsky* Ariel Boyarsky, Hongseok Namkoong, Jean Pouget-Abadie,

Experiments on online marketplaces and social networks suffer from interference, where the outcome of a unit is impacted by the treatment status of other units. We propose a framework for modeling interference using a ubiquitous deployment mechanism for experiments, staggered roll-out designs, which slowly increase the fraction of units exposed to the treatment to mitigate any unanticipated adverse side effects. Our main idea is to leverage the temporal variations in treatment assignments introduced by roll-outs to model the interference structure. Since there are often multiple competing models of interference in practice we first develop a model selection method that evaluates models based on their ability to explain outcome variation observed along the roll-out. Through simulations, we show that our heuristic model selection method, Leave-One-Period-Out, outperforms other baselines. Next, we present a set of model identification conditions under which the estimation of common estimands is possible and show how these conditions are aided by roll-out designs. We conclude with a set of considerations, robustness checks, and potential limitations for practitioners wishing to use our framework.

Causal Inference and SUTVA/Consistencies Violations

Nonparametric Causal Survival Analysis under Clustered Interference Chanhwa Lee*
Chanhwa Lee, Michael Hudgens,

Interference arises when a unit’s treatment affects the outcome of other units. Sometimes, units are grouped into clusters, where it is reasonable to assume interference only occurs within cluster, i.e., clustered interference. Several methods exist for estimating various causal estimands under clustered interference from observational data, but either (i) the estimands lack real-world relevance, (ii) the estimators rely on parametric models, and/or (iii) the methods do not accommodate right-censored outcomes. To address these issues, we introduce a general framework for estimating treatment effects in the presence of clustered interference and right censoring. Our method is applicable
to any stochastic policy which modifies the propensity score distribution and thus relevant across diverse settings. Nonparametric sample splitting estimators are constructed, allowing for flexible data-adaptive estimation of nuisance functions, and are consistent and asymptotically normal, converging at the usual parametric rate. Simulation studies demonstrate the finite sample performance of the proposed estimators, and the method is applied to a cholera vaccine study in Bangladesh.

Causal Inference and SUTVA/Consistencies Violations

Estimating Higher-order Spillover Effects on Network Qixiang Xu* Qixiang Xu, Laura Forastiere,

“Interference, where a unit’s outcome is affected by the other units’ treatments through network connections, is a phenomenon of interest. Researchers often focus on the spillover effect from first-order neighbors. However, the prevailing approach often involves the neighborhood interference assumption, which can be restrictive. This paper proposes a broader assumption, the generalized interference assumption, which allows potential outcomes to be influenced by a wider range of networks, referred to as the ’interference set’. This might include a community detected through an algorithm, or units that can be reached through a finite network path. We define new causal estimands to quantify spillover effects from units at a specific network distance h. We employ two hypothetical Bernoulli distributions with different probabilities for the h-order neighborhood for the rest of the units in the interference set. We first derive the bias of a commonly used ‘naive’ estimator which relies on a wrong interference set or incorrect exposure mapping functions. We then develop new Horvitz-Thomson and Hajek estimators and corresponding weighted regression estimators under this broader assumption. We assess the bias of naive estimators and the performance of our estimators in different interference scenarios and random graphs through simulations. Finally, we apply these estimators in a two-stage randomized trial in Honduras, evaluating a maternal and child health intervention.

Causal Inference and SUTVA/Consistencies Violations

Operational Challenges in Scaling Randomized Trials: The Role of Capacity Constraints
Hannah Li* Hannah Li, Justin Boutilier, Jonas Jonasson,

A concern about service interventions, commonly found in domains like public health and education, is that promising interventions at the randomized controlled trial (RCT) stage may not perform well at scale. Although many factors contribute to this difficulty in scaling, in this work we highlight and isolate the effects of an operational factor: capacity constraints.
If an intervention requires a service that is capacity constrained, then participants in the RCT may face a waiting time that depends on the number of providers and number of other participants in the system. We show that this dependency may lead to violations of the Stable Unit Treatment Value Assumption (SUTVA) and creates scaling issues.

We consider a case study of a mobile health platform designed to improve patients' adherence to tuberculosis treatment. By modeling patients' interactions as a queueing system, we demonstrate that the effects observed in an RCT may decrease when scaling up to a larger patient population, due to the system's limited capacity.

Furthermore, we find a counterintuitive implication for conventional power analysis: increasing the sample size of an RCT without appropriately expanding capacity can paradoxically decrease the study's power. To address this, we leverage principles from operations research and introduce a method for joint power and capacity analysis that leverages the underlying structure of these interventions in order to increase power.

Causal Inference and SUTVA/Consistencies Violations

Causal inference with spatio-temporal data using process-informed stochastic interventions Nathan Wikle* Nathan Wikle, Corwin Zigler,

Causal inference with spatio-temporal data is often challenging due to the presence of interference: outcomes for observational units depend on some combination of local and non-local treatment. This is especially relevant when an individual's treatment exposure is determined, in part, by some underlying physical process. For example, air pollution exposure is a function of both the locations of emissions sources (e.g., power plants, roads, etc.) as well as the physical-chemical process governing pollution transport. In this talk, we propose causal estimands that are defined with respect to stochastic interventions informed by the physical process; importantly, these estimands can accommodate both interference and positivity violations. In particular, we estimate the expected change in the number of outcome events in a specific area under different stochastic exposure distributions, where the stochastic distributions correspond to counterfactual exposure levels simulated from a dynamic spatio-temporal model of the physical process. We develop an augmented inverse probability of treatment weighting estimator for spatio-temporal data with a desirable double robustness property, and propose methods to assess its sensitivity to unmeasured confounding, positivity violations, and uncertainty in the physical process model. Finally, we use the proposed methods to estimate the expected change in pediatric asthma rates in Texas under two competing air pollution emissions scenarios.

Causal Inference Education

The Interpretation of Associational Language in Research Statements Noah Stovitz* Noah Stovitz, Ian Shrier, Jake Quilty-Dunn, Jennifer Hill,

Researchers often use ambiguous language, and even clearly written non-causal language may be misinterpreted as causal. The purpose of this study is to evaluate the effect of language on the interpretation of causality in research articles. We surveyed under- or recently graduated university students. We randomized them to see one of 5 different “linking words” between variables that might represents causes and outcomes (ordered by perceived level of causal implication: affects, increases, predicts, increased with, correlated with) within three different contexts which varied by a priori level of perceived causal relationship (exercise and dehydration: likely causal, study abroad and graduation: possibly causal or non-causal, born early in week and intelligence: likely non-causal). In the first 60 respondents with complete data, the proportion of respondents who reported the claim as having a strong causal implication was only 25% for affects (our a priori strongest causal implication), 53% for increases, 35% for predicts, 38% for increased with, and 22% for correlated with. The context affected the interpretation for 45% (27/60) of the participants. Among those participants, 19% (5/27) interpreted the claim as being more causal even though the context shifted from likely causal to likely non-causal. We conclude that interpretations of linking words are complex, with the likelihood of a true causal effect affecting the interpretation.

Causal Inference Education

Towards Accessible Proof Techniques for Non-Identification In Causal Inference Juan Mendez* Juan Mendez, Hadassah Lurbur, Rohit Bhattacharya,

Increasingly causal inference is becoming a staple in undergraduate and graduate curricula in fields such as computer science, (bio)statistics, economics, and epidemiology. A fundamental point that must be emphasized in these courses is why certain causal parameters cannot be computed from observed data without sufficient restrictions on the data generating process. Explanations of the non-identifiability of causal parameters, even at the graduate level, often only provide a rough intuition, or rely on specially constructed counter examples that do not generalize well if the student were interested in examining different parameters.

We explore classic identification results in causal inference and work through proofs of identification and non-identification using a general and (we hope) more accessible framework that uses concepts a student might encounter in undergraduate computer science or statistics courses-–concepts in discrete mathematics like probability distributions, bijections, and proof by contradiction—and an introductory course in probabilistic graphical models for more advanced proofs. We first present arguments of non-identification of objects like cross-world parameters (joint distributions over potential outcomes that disagree on the value of treatment assignment) and build up to other classical results in causal inference, such as necessary and sufficient conditions for identifying the effect of one variable on all other variables (Tian & Pearl, 2002).

Causal Inference in Networks

Estimating heterogenous spillover effects on network neighbors to identify influential and susceptible individuals Yihan Bao* Yihan Bao, Laura Forastiere,

Due to peer influence, behavioral interventions received by a unit are likely to affect the behavioral outcomes of other socially connected units. Under interference, spillover effects have been defined in previous works by contrasting potential outcomes under a different number of treated units or under different treatment allocations in the interference set. In our work, under a partial interference assumption, we define average and conditional spillover effects of having a network neighbor treated vs not treated, while the treatment of other units in the interference set is randomly assigned under a given allocation strategy. By varying the conditioning sets, we can assess the heterogeneity with respect to the characteristics of the influencer and those of the influencee, to identify influential and susceptible individuals, respectively. Under a super-population perspective, we develop IPW estimators for average and heterogeneous influence effects, with marginal structural models for continuous covariates. We then use our estimators to investigate the characteristics of influencers and susceptible individuals in a two-stage randomized study conducted in Honduras to assess the spillover effects of a behavioral intervention. Here, we further address the presence of non-compliance in the second stage by replacing the theoretical second-stage treatment probability with the estimated propensity score, conditional on the first stage.

Causal Inference in Networks

Quasi-randomization tests for network interference Supriya Tiwari* Supriya Tiwari, Pallavi Basu,

Many classical inferential approaches fail to hold when interference exists among the population units. This amounts to the treatment status of one unit affecting the potential outcome of other units in the population. Testing for such spillover effects in this setting makes the null hypothesis non-sharp. There is a growing body of literature that considers this problem in an experimental setup with the network structure amongst the population to be fixed and assumed to be given. An interesting approach to tackling the non-sharp nature of the null hypothesis in this setup is constructing conditional randomization tests such that the null is sharp on the restricted population. In randomized experiments, conditional randomized tests hold finite sample validity. However, such approaches are computationally intensive as finding these appropriate sub-populations can involve solving an NP-hard problem. In this paper, we view the network amongst the population as a random variable instead of fixed and treat the given network as the observed outcome of the network random variable. We propose a new approach that builds a conditional quasi-randomization test. We highlight that the approach is easier to implement than the current state-of-the-art methods. We conduct a simulation study to verify the finite-sample validity of our approach and illustrate our methodology to test for interference in a weather insurance adoption experiment run in rural China.

Causal Inference in Networks

Average and Conditional Inward and Outward Spillovers of One Unit's Treatment under Network Interference Fei Fang* Fei Fang, Laura Forastiere,

In a connected social network, users may have varying levels of influence on others when they themselves receive interventions. For example, giving an advertisement to a more influential person can have on average a greater impact on others' purchase decisions. Understanding and evaluating these effects can provide valuable insights for various applications such as targeting strategies in marketing and behavioral interventions in public health. Under a partial interference assumption, we define influence effects in two ways: i) the inward average spillover effect on a unit's outcome of a neighbor's treatment, and ii) the outward average spillover of a unit's treatment on their neighbors' outcomes. We investigate the comparison between the two causal effects in directed networks with different properties, including the conditions under which they are equivalent. Additionally, we develop Horvitz-Thompson estimators for assessing both effects, on average and conditioning on categorical covariates, as well as weighted least square estimators for these effects conditioning on continuous covariates. We derive design-based variance estimators and establish the consistency and asymptotic normality. Through simulations, we verify the empirical performance of our proposed estimators. Finally, we employ our approach to investigate inward and outward average and conditional spillover effects of an information session on the adoption of weather insurance among rice farmers in China.

Causal Inference in Networks

Data-adaptive exposure thresholds for the Horvitz-Thompson estimator of the Average Treatment Effect in experiments with network interference Vydhourie Thiyageswaran* Vydhourie Thiyageswaran, Jennifer Brennan,

Randomized controlled trials on network data often suffer from interference, a SUTVA violation in which a unit's treatment assignment affects the outcomes of its neighbors. A popular method to reduce the bias, caused by interference, in estimating the Average Treatment Effect (ATE) is to apply the Horvitz-Thompson estimator of the ATE with an exposure mapping: a function that identifies which units in a given randomization are not subject to interference. For example, an exposure mapping may specify that any unit with at least X% of its neighbors having its same treatment status does not experience interference. In this work we propose a data-adaptive method to select this "X%" threshold, which greatly affects the mean squared error of the Horvitz-Thompson estimator but is often difficult to elicit from domain experts. Our method estimates the bias and variance of the Horvitz-Thompson estimator under different thresholds using a linear dose-response model of the potential outcomes. We present simulations illustrating that our method improves upon non-adaptive choices of the threshold for cycle graphs (and their 2k-degree extensions). Furthermore, we demonstrate that our method is robust to deviations from the linear potential outcomes model.

Causal Inference in Networks

A Bayesian approach to Estimate Causal Peer Influence Using Latent Location for Unmeasured Confounding of Homophily Seungha Um* Seungha Um, Samrachana Adhikari,

Researchers have been focused on estimating causal inferences to understand how individual's behavior is influenced by the behaviors of their peers in observational studies on social networks. Identifying and estimating the peer influence, however, is challenging due to frequent confounding with homophily, where people tend to connect with those who share similar characteristics with them. Moreover, as the attributes driving homophily are generally not directly observed and serve as unobserved confounders, the identification and estimation of causal peer influence is not possible. In this paper, we address this challenge by leveraging latent locations inferred from the network itself to disentangle homophily from causal peer influence, and extend this approach to multiple networks by adopting a Bayesian hierarchical modeling framework. To model nonlinear response surfaces capturing the effective range of peer influence, we employ a Bayesian nonparametric model, specifically Bayesian Additive Regression Trees (BART). We propose an integrated Bayesian framework to account for the uncertainty in inferring latent locations and outcome modeling. We establish a nonparametric identification of causal peer influence in the presence of unmeasured network confounding without imposing fany parametric restrictions on the outcome model. To illustrate the applicability of our method in estimating causal peer influence, we utilize both simulated data and advice-seeking networks real data.

Causal Inference in Networks

Nonparametric Network Causal Inference for Continuous Exposures in Mobile Source Air Pollution Salvador Balkus* Salvador Balkus, Nima Hejazi, Rachel Nethery, Scott Delaney,

Continuous exposures pose a challenge for traditional causal inference estimators: positivity is frequently violated, and the intervention of “setting every unit’s exposure to exactly X” produces unrealistic counterfactuals. Modified Treatment Policies (MTPs) – interventions that depend on each unit’s naturally observed exposure – resolve these issues. But what if the outcome of each unit depended not just on its own exposure, but also on the exposure of other units in a network? For example, in observational mobile source air pollution studies, the pollution in a given region depends not only on vehicles registered within it but also on vehicles that commute in from other regions. Our work connects recent theory in nonparametric causal inference under general network interference to MTPs. We review the mathematical construction of candidate estimators, compare multiple variance estimation strategies, and introduce an open-source software implementation in Julia. Estimators are evaluated under various simulation settings and applied to a mobile source air pollution case study.

Causal Inference in Networks

Causal Inference With Contagion and Latent Homophily Under Full Interference Yufeng Wu* Yufeng Wu, Rohit Bhattacharya,

Dependent data poses a serious challenge for valid causal inference. In some cases, we observe a single realization of a network of individuals, all of whom may depend on each other – a setting termed “full interference.” Tchetgen Tchetgen et al (2021) developed the auto-g computation method for computing network causal effects in this setting. Their method assumes that all relevant confounders are observed and that there is no dependence induced by unmeasured common causes between individuals. In other words, the method allows for contagion and interference effects, but assumes the absence of latent homophily. In this work, we propose a nonparametric test that can be used to distinguish between dependence due to contagion and latent homophily. This test acts as a verification tool for the auto-g computation method, providing a way to accept or reject its model assumptions. In cases where there is dependence due to latent homophily, the auto-g method produces biased estimates of network causal effects, so we propose a modified identification and estimation strategy for settings where dependence could be induced due to either latent homophily or contagion. We evaluate the effectiveness of our method through simulation studies and a real-world data application on social networks.

Eric J Tchetgen Tchetgen, Isabel R Fulcher, and Ilya Shpitser. Auto-g-computation of causal effects on a network. Journal of the American Statistical Association, 116(534): 833–844, 2021

Causal Inference in Networks

Staggered Rollout Designs with Clustering Mayleen Cortez-Rodriguez* Mayleen Cortez-Rodriguez, Matthew Eichhorn, Christina Lee Yu,

Many approaches for estimating causal effects under interference rely on total knowledge of the underlying causal network, which is often unrealistic in practice. Recent work has shown that even with no network knowledge, one can still obtain unbiased estimates for causal effects by leveraging a staggered rollout experimental design and polynomial interpolation (PI). This approach can have high variance due to extrapolating a polynomial far from the support of the data. Additionally, it disregards potentially useful information or covariate data that may be available about the graph.

In this work, we investigate PI estimators under a two-stage experimental design wherein a graph clustering in the first stage selects a subpopulation on which a staggered rollout design is implemented in the second stage. Limiting the experiment to this subpopulation allows for a larger experimental budget in the second stage of the experiment, reducing the extrapolation error in high degree models. However, this approach can increase sampling error as the chosen subset may not be representative of the entire population. We provide experiments that illustrate the impact of homophily and clustering quality on this tradeoff between extrapolation and sampling error. We also explore the robustness of PI estimators under model misspecification. These experiments help us understand when clustering is a good choice for this style of estimator and design.

Difference in Differences

Exploring the Impact of Health IT Features on Quality of Hospital Care Raluca Cobzaru*
Raluca Cobzaru, Roy Welsch, Stan Finkelstein, Zach Shahn,

As policies with staggered adoption – such as Medicaid expansion across states – have been gaining traction in the public health sector, so has the topic of difference-in-differences (DiD) methods which allow for time-varying treatment strategies. We utilize a novel adjustment framework for time-dependent covariates in the DiD setting to study the causal effect of implementing certain EHR system features (e.g. clinical decision support, care coordination across departments, automated drug interaction checks, etc.) on hospital performance metrics (e.g. 30-day mortality and readmission rates) over time. Our results help inform Health IT adoption policies and ease the operational burden on hospitals by identifying a sufficient set of EHR features associated with improved quality of care.

Difference in Differences

Sequential Synthetic Difference in Differences Aleksei Samkov* Aleksei Samkov, Dmitry Arkhangelsky,

We study the estimation of treatment effects of a binary policy in environments where the rollout of the treatment is staggered. We show that the identification problem for a particular factor model can be solved using the Synthetic Difference in Difference (SDiD) method. We analyze the class of the corresponding estimators and connect their asymptotic behavior to a class of oracle estimators. We derive the optimal oracle estimator in this class and use this connection to propose a new procedure — sequential SDiD — and show that it is asymptotically unbiased, normal, and efficient. The method developed in this paper presents a natural alternative to the conventional DiD strategies in staggered adoption designs.

Generalizability/Transportability

Paradoxes and resolutions for semiparametric fusion of individual and summary data
Wenjie Hu* Wenjie Hu, Ruoyu Wang, Wei Li, Wang Miao,

Suppose we have available individual data from an internal study and various types of summary statistics from relevant external studies. External summary statistics have been used as constraints on the internal data distribution, which promised to improve the statistical inference in the internal data; however, the additional use of external summary data may lead to paradoxical results: efficiency loss may occur if the uncertainty of summary statistics is not negligible and a large estimation bias can emerge even if the bias of external summary statistics is small. We investigate these paradoxical results in a semiparametric framework. We establish the semiparametric efficiency bound for estimating a general functional of the internal data distribution, which is shown to be no larger than that using only internal data. We propose a data-fused efficient estimator that achieves this bound so that the efficiency paradox is resolved. Besides, we propose a debiased estimator that can achieve the same asymptotic distribution as the oracle estimator as if one knew whether the summary statistics were biased or not. Simulations and
application to a Helicobacter pylori infection dataset are used to illustrate the proposed methods.

Generalizability/Transportability

Recovering target causal effects from post-exposure selection induced by missing outcome data Johan de Aguas* Johan de Aguas, Johan Pensar, Tomás Varnet, Guido Biele,

Two significant challenges to the validity of causal claims are confounding bias and selection bias. The latter can arise through informative missingness, where partial information about units in the population is missing, censored, or coarsened due to factors related to the exposure, the outcome, or their consequences. We extend existing graphical criteria to address selection bias induced by missing outcome data by leveraging post-exposure variables. We introduce the generalized adjustment criteria with post-exposure variables (GACPE), which supports a recovered causal effect based on sequential regressions. A refined estimator is further developed by applying targeted minimum loss estimation (TMLE). Under certain regularity conditions, this estimator is multiply-robust, ensuring consistency even in scenarios where inverse probability weighting (IPW) and the sequential regressions approach fall short. A simulation study with various scenarios contrasts the relative robustness and efficiency of the two proposed solutions against other classical estimators. As a motivating application case, we study the effects of pharmacological treatment for attention-deficit/hyperactivity disorder (ADHD) upon the scores of national tests taken by diagnosed Norwegian schoolchildren. Findings support the accumulated clinical evidence affirming a positive but small effect of stimulant medication on school performance.

Generalizability/Transportability

Estimating Causal Effects with Error-Prone Exposures Using Control Variates Keith Barnatchez* Keith Barnatchez, Kevin Josey, Rachel Nethery, Giovanni Parmigiani,

Exposure measurement error poses a common, yet often ignored, challenge to performing causal inference in observational studies. Existing methods accounting for exposure measurement error largely rely on restrictive parametric assumptions for not only the measurement error mechanism, but also the outcome and exposure models used to estimate a causal effect. There remains a critical need for assumption-lean estimation methods that can flexibly accommodate different study designs while possessing desirable theoretical properties. In this paper, we address these needs by proposing estimators based on the control variates framework of Yang and Ding (2020). Drawing connections between the measurement error, generalizability and transportability, and missing data literatures, we show that our approach can be implemented in various two-stage study designs–where one obtains gold-standard exposure measurements for a small subset of the initial study sample–to address biases induced by measurement error for estimating general causal quantities. Under standard causal inference assumptions, our method inherits desirable double-robustness properties, including scenarios where the two-stage sampling probabilities are unknown. Through simulation studies, we show our approach performs favorably to leading methods under various two-stage sampling schemes. Finally, we test our method on observational electronic health record data on HIV outcomes from the Vanderbilt Comprehensive Care Clinic.

Generalizability/Transportability

Selective Experimentation and Stability Radius Models Adam Bouyamourn* Adam Bouyamourn,

Choosing an internally-valid causal inference design is not sufficient to ensure that the conclusions of a paper are true: researchers may strategically select experiments whose conclusions do not match the inferences that would have been drawn if a larger or more representative experiment was conducted. Using a formal model to study researcher incentives, I first show that the problem is one of information revelation: the problem is that honest researchers cannot credibly communicate that they did not strategically select their experiment. Drawing inspiration from the control theory literature on stability radius models, I then show that reporting a stability radius can induce sufficient information disclosure to allow an audience to assess whether or not to trust the conclusion of a given paper. Then, I develop two empirical tools for estimating stability radii, each using a conformal wrapper for inference: the first using Support Vector Machines, the second using a factor model.

Generalizability/Transportability

Revisiting Representativeness Haidong Lu* Haidong Lu,

The concept of representativeness has long been a key concept in epidemiology and other social sciences. Nonetheless, the interpretations of what constitutes representativeness and non-representativeness can differ among researchers. Some associate non-representativeness with a lack of generalizability, while others tend to conflate it with collider bias. In this paper, I aim to provide a comprehensive understanding of the underlying mechanisms of representativeness. Specifically, I will introduce and delineate two fundamental concepts: sample representativeness and estimate representativeness. Through the use of two distinct simulation studies involving causal diagrams, I will elucidate the reasons behind the occurrence of estimate non-representativeness when there is sample non-representativeness. Furthermore, I will discuss the settings in which estimate representativeness can be maintained even in the presence of sample non-representativeness.

Generalizability/Transportability

Robust Transfer Learning Between 2020 and 2024 U.S. Presidential Elections: A Case Study of Transporting Inference from a 2020 Digital Advertising Campaign Xinran Miao*
Xinran Miao, Hyunseung Kang,

In 2020, Aggarwal et al conducted a 2 million-person randomized experiment analyzing the impact of online advertising against Donald Trump on voter turnout in five battleground states. While the estimated overall effect of the online campaign was “effectively equivalent to zero”, this large experiment may offer important insights about the 2024 campaign. To this end, we propose a simple estimation procedure that transports average treatment effect (ATE) from a source experimental study to a target population. Our procedure does not necessarily rely on the transportability assumption, which states that the source and target populations only differ by a set of common covariates and cannot be verified by data. Instead, we use a sensitivity analysis approach where we provide a range of plausible estimates of the ATE given different degrees of violation of the assumption. Applying the proposed framework, we transport the ATE of digital advertising on registered voters to battleground states in 2024 and we find that even a tiny change from 2020 (1.01 in the turnout odds ratio among the treated) will result in a statistically significant effect of the digital advertising on turnout in Georgia, Michigan, Wisconsin, and Nevada.

Generalizability/Transportability

Super-Efficient Estimation of Average Treatment Effect based on Randomized Controlled Trial Augmented with External Controls or Observational Study Sky Qiu* Sky Qiu, Lars van der Laan, Mark van der Laan,

We consider the problem of estimating the average treatment effect (ATE) when both randomized control trial (RCT) data and real-world data (RWD) are available. We decompose the ATE estimand as the difference between a pooled-ATE estimand that integrates RCT and RWD and a bias estimand that captures the RCT study indicator's conditional effect on the outcome. We introduce an adaptive targeted minimum loss-based estimation (A-TMLE) framework to estimate them. We prove that the A-TMLE estimator is root-n-consistent and asymptotically normal. Moreover, it achieves the super-efficiency one would obtain had one known the oracle model for the conditional effect of the study indicator on the outcome. Consequently, the smaller the working model of the bias induced by the RWD is, the greater our estimator's efficiency, while our estimator will always be at least as efficient as an efficient estimator that uses the RCT data only. We demonstrate our estimator's superior performance through simulations. We also apply our method to the DEVOTE trial, with RWD from Optum. A-TMLE helps utilize RWD to improve the efficiency of randomized trial results without biasing the estimates of intervention effects. This approach could allow for smaller, faster trials, decreasing the time until patients can receive effective interventions.

Heterogeneous Treatment Effects

Principal Stratification with Continuous Post-Treatment Variables: Nonparametric Identification and Semiparametric Estimation Sizhu Lu* Sizhu Lu,

Post-treatment variables often complicate causal inference. They appear in many scientific problems, including noncompliance, truncation by death, mediation, and surrogate endpoint evaluation. Principal stratification is a strategy that adjusts for the potential values of the post-treatment variables, defined as the principal strata. It allows for characterizing treatment effect heterogeneity across principal strata and unveiling the mechanism of the treatment on the outcome related to post-treatment variables. However, the existing literature has primarily focused on binary post-treatment variables, leaving the case with continuous post-treatment variables largely unexplored, due to the complexity of infinitely many principal strata that challenge both the identification and estimation of causal effects. We fill this gap by providing nonparametric identification and semiparametric estimation theory for principal stratification with continuous post-treatment variables. We propose to use working models to approximate the underlying causal effect surfaces and derive the efficient influence functions of the corresponding model parameters. Based on the theory, we construct doubly robust estimators and implement them in an R package.

Heterogeneous Treatment Effects

Causal Q-Aggregation for CATE Model Selection Hui Lan* Hui Lan, Vasilis Syrgkanis Syrgkanis,

Accurate estimation of conditional average treatment effects (CATE) is at the core of personalized decision making. While there is a plethora of models for CATE estimation, model selection is a nontrivial task, due to the fundamental problem of causal inference. Recent empirical work provides evidence in favor of proxy loss metrics with double robust properties and in favor of model ensembling. However, theoretical understanding is lacking. Direct application of prior theoretical work leads to suboptimal oracle model selection rates due to the non-convexity of the model selection problem. We provide regret rates for the major existing CATE ensembling approaches and propose a new CATE model ensembling approach based on Q-aggregation using the doubly robust loss. Our main result shows that causal Q-aggregation achieves statistically optimal oracle model selection regret rates of log(M)/n (with M models and n samples), with the addition of higher-order estimation error terms related to products of errors in the nuisance functions. Crucially, our regret rate does not require that any of the candidate CATE models be close to the truth. We validate our new method on many semi-synthetic datasets and also provide extensions of our work to CATE model selection with instrumental variables and unobserved confounding.

Heterogeneous Treatment Effects

Estimating Heterogeneous Treatment Effects for General Responses Zijun Gao* Zijun Gao, Trevor Hastie,

Heterogeneous treatment effect models allow us to compare treatments at subgroup and individual levels, and are of increasing popularity in applications like personalized medicine, advertising, and education. In this work, we first survey different causal estimands used in practice, which focus on estimating the difference in conditional means. We then propose DINA — the difference in natural parameters — to quantify heterogeneous treatment effect in exponential families and the Cox model. For binary outcomes and survival times, DINA is both convenient and more practical for modeling the influence of covariates on the treatment effect. Second, we introduce a meta-algorithm for DINA, which allows practitioners to use powerful off-the-shelf machine learning tools for the estimation of nuisance functions, and which is also statistically robust to errors in inaccurate nuisance function estimation. We demonstrate the efficacy of our method combined with various machine learning base-learners on simulated and real datasets.

Heterogeneous Treatment Effects

Causal machine learning for heterogeneous treatment effects in the presence of missing outcome data Matthew Pryce* Matthew Pryce, Karla Diaz-Ordaz, Stijn Vansteelant, Ruth Keogh,

In recent years, there has been a growing interest in exploring personalized treatment/policy decisions. However, estimating heterogenous treatment effects often requires the use of large, rich datasets, containing either high dimensional data, or complex correlations between variables. As a result, causal machine learning estimators, such as the DR-learner have grown in popularity, offering flexible and efficient tools for exploring heterogeneity.
Additionally, data often contains loss of follow up, with outcomes missing. To handle this, researchers often run an imputation model, or use inverse-probability censoring weights, reducing the robustness and efficiency of the CATE estimation process. Therefore, we propose an extension of the DR-learner which handles missing outcome data through an orthogonal implementation of inverse censoring weights. Our robust solution assumes missing at random outcome data and utilizes semi-parametric theory to derive an estimator for the CATE via a two-step process: debiasing outcome predictions via the EIF of the average treatment effect; then running a pseudo-outcome regression to obtain estimates of the CATE.
We demonstrate the utility of the approach mathematically, providing excess risk bounds, and empirically, through a simulation study and data example. We also provide a debiased MSE validation metric for the CATE, along with an extension of the approach which allows for time-varying drivers of missingness to be adjusted for.

Heterogeneous Treatment Effects

A Comparison of Causal Forests and the DR-Learner for Estimating Conditional Average Treatment Effects Qi Zhang* Qi Zhang, Ya-Hui Yu, Ashley Naimi,

Conditional average treatment effects (CATEs) hold great promise for precision medicine, particularly in settings where effect modification is likely. Theoretical work has developed methods to estimate CATEs, including the double-robust (DR) learner and the causal forest algorithm. Here, we conduct a simulation study comparing the finite sample properties of the DR learner and the causal forest algorithm. We explore performance in a range of scenarios with a binary effect modifier and when a set of conditioning variables are included with varying degrees of effect modifiers present. Scenarios explored different effect parametrizations, sample sizes, proportions of modifying to non-modifying variables, and number of confounding variables. For all analyses, we used 10-fold cross fitting, and linear projection approach to identify pre-specified modifiers. Our preliminary results suggest that both the causal forest and the DR learner have good 95% confidence interval coverage in most settings. However, the DR learner outperformed the causal forest in coverage (93% vs. 88%) under the scenarios of strong treatment effect but low heterogeneity. We also found that the best linear projections may not always reliably identify pre-specified effect modifiers when either method is used, especially in small sample sizes (with a successful identification of 53% at best scenario). This will provide practical insights to guide method selection for estimating CATEs in empirical research.

Heterogeneous Treatment Effects

A Differential Effect Approach to Partial Identification of Treatment Effects Kan Chen* Kan Chen, Bingkai Wang, Dylan Small,

We consider identification and inference for the average treatment effect and heterogeneous treatment effect conditional on observable covariates in the presence of unmeasured confounding. Since point identification of these treatment effects is not achievable without strong assumptions, we obtain bounds on these treatment effects by leveraging differential effects, a tool that allows for using a second treatment to learn the effect of the first treatment. The differential effect is the effect of using one treatment in lieu of the other. We provide conditions under which differential treatment effects can be used to point identify or partially identify treatment effects. Under these conditions, we develop a flexible and easy-to-implement semi-parametric framework to estimate bounds and leverage a two-stage approach to conduct statistical inference on effects of interest. The proposed method is examined through a simulation study and a case study that investigates the effect of smoking on the blood level of cadmium using the National Health and Nutrition Examination Survey.

Heterogeneous Treatment Effects

A nonparametric Gail-Simon test and estimand for qualitative effect heterogeneity Aaron Hudson* Aaron Hudson, Mats J. Stensrud, Oliver Dukes, Ricardo Brioschi,

Qualitative heterogeneity or effect modification, occur when treatment is beneficial for certain subgroups and harmful for others. This specific type of heterogeneity is of clinical interest when treatment decisions will be tailored to individual characteristics. The problem of testing for qualitative heterogeneity has been well-studied when the comparison is made between finite subgroups; for example, Gail and Simon (1985) proposed a likelihood ratio test in the context of discrete covariates. However, the problem is more challenging when the potential effect modifiers are continuous, and one wishes to infer the conditional average treatment effect under a nonparametric model. In this talk, we propose a class of nonparametric tests for qualitative heterogeneity as a natural extension of the Gail-Simon test. Compared with some recent approaches, our proposal can incorporate a variety of structured assumptions on the conditional average treatment effect, extends to moderate/high-dimensional covariates and does not require sample splitting. The utility of the proposal is borne out in simulation studies and a re-analysis of a recent clinical trial.

Heterogeneous Treatment Effects

Revisiting a problem of Kolmogorov with application to individual treatment effects Zhehao Zhang* Zhehao Zhang, Thomas Richardson,

We revisit the following problem, proposed by Kolmogorov: given prescribed marginal distributions F and G for random variables X and Y respectively, characterize the set of compatible distribution functions for the sum Z=X+Y. Bounds on the distribution function for Z were given by Makarov (1982), and Frank, Nelson and Schweizer (1987), the latter using copula theory. However, though they obtain the same bounds, they make different assertions concerning their sharpness. In addition, their solutions left some open problems in the case when the given marginal distribution functions are discontinuous. These issues have led to some confusion and erroneous statements in subsequent literature, which we correct.

Kolmogorov's problem is closely related to inferring possible distributions for individual treatment effects Y(x=1) – Y(x=0) given the marginal distributions Y(x=0) and Y(x=1); the latter being identified from a randomized experiment. We use our new insights to sharpen results due to Fan and Park (2010) concerning individual treatment effects, and to fill some other logical gaps.

Heterogeneous Treatment Effects

Minimax Optimal Estimates of Individual Causal Effects in Panel Data under Heterogeneous Two Way Fixed Effects Models Calvin Tolbert* Calvin Tolbert, Christina Lee Yu, Xumei Xi, Yudong Chen,

Consider estimating individual causal effects of a treatment on an individual i at time t from panel data assuming the two-way fixed effects model, where the outcome under treatment and control can be written as a sum of a unit-specific effect and a time-specific effect.
For any given observation pattern specifying the locations of observed entries, we establish error upper bounds customized to each entry as a function of the observation pattern, along with minimax lower bounds on the entry specific squared error that match the upper bounds to a factor of 4. Our results show that the minimax achievable squared error for recovery of the causal effect associated to the (i,t) unit-time pair is proportional to the effective resistance of (i,t) in the bipartite graph associated to the observation pattern. Pairs (i,t) with low effective resistance are well connected in the graph, resulting in a high minimax error bound.
The estimator that achieves this performance is a weighted sum of the observations where the weights are derived from the unit (i,t) electrical flow. The algorithm recovers individual causal effects and the unit-specific and time-specific confounders up to the achievable precision.
Our results provide necessary and sufficient conditions for identifiability of individual causal effects, along with optimally efficient fine-grained entrywise estimates.

Instrumental Variables

Identifying Causal Effects of Nonbinary, Ordered Treatments using Multiple Instrumental Variables Nadja van ‘t Hoff* Nadja van ‘t Hoff,

This paper addresses the challenge of identifying causal effects of nonbinary, ordered treatments with multiple binary instruments. Next to presenting novel insights into the widely-applied two-stage least squares estimand, I show that a weighted average of local average treatment effects for combined complier populations is identified under the limited monotonicity assumption. This novel causal parameter has an intuitive interpretation, offering an appealing alternative to two-stage least squares. I employ recent advances in causal machine learning for estimation. I further demonstrate how causal forests can be used to detect local violations of the underlying limited monotonicity assumption. The methodology is applied to study the impact of community nurseries on child health outcomes.

Instrumental Variables

Online Training, Working from Home, and Industry Productivity Octavio M. Aguilar* Octavio M. Aguilar,

In this paper, I use an instrumented difference-in-differences approach to investigate the impact of online training on industry productivity both during and after the COVID-19 Recession. I provide new evidence indicating that industries with higher exposure to online training experience a substantial and enduring decline in productivity —1.3% in 2021 and 5.2% in 2022. To understand the mechanisms behind this productivity decrease, I investigate the positive relationship between working from home (WFH) and online training. Additionally, I explore the time allocation of employees to online training post-COVID-19 Recession. I find the following: (1) for every 1% increase in WFH, there is a 4.5% increase in online training, and (2) increases in WFH contribute to, on average, an additional 130 minutes per-day spent on online training. I argue that the decline in productivity can be attributed to the reduced supervision of employees working remotely. Consequently, employees allocate labor hours to enhance their human capital.

Instrumental Variables

Impact of Prosecutorial Systems on Charging Decisions: Application of the Instrumental Variable Method Takuma Iwasaki* Takuma Iwasaki,

In the field of criminal law and policy, there has been limited research applying causal inference methods, because most new laws and systems apply to all entities uniformly and simultaneously in order to maintain equality, which makes it challenging to establish a control group. Focusing on a case where two comparable systems coexist, my research shows the impact of organizational systems on prosecutors' charging rates.
In most countries, prosecutors' offices are organized using one of two systems. In the first system, the same prosecutor handles cases all the way from investigation to trial. In the second system, different prosecutors are responsible for investigation and trial, respectively, where one prosecutor decides to charge, but does not have to spend its time on the subsequent trial.
Interestingly, both systems coexist in Japan, but there are unobserved confounders, such as severity of criminal cases in each area, affecting both the system selection and charging rates. However, the system selection primarily hinges on geographical proximity to major prosecutor's offices, making the proximity an instrument, since it does not directly impact charge rates, and is unrelated to the unobserved confounders.
My research reveals that the second system increases charge rates by 15.6% [95% CI: 2.4 – 28.8%], suggesting that prosecutors in the second system charge more aggressively since they do not have to "pay the price" for it.

Instrumental Variables

Robust, sharp bounds for principal effects in randomized encouragement design with an ordinal measure of compliance Xiaobin Zhou* Xiaobin Zhou, Chan Park, Hyunseung Kang,

Principal stratification is a popular framework to study treatment effects in subpopulations defined by post-treatment variables (i.e. principal strata). For example, in randomized encouragement design, principal stratification can be used to define treatment effects among people with different levels of compliance. However, point identification of principal strata effects requires stringent identifying assumptions (e.g. exclusion restriction, principal ignorability), especially when the post-treatment variable is not binary. In this work, we establish nonparametric, sharp bounds for treatment effects with covariate-adjusted linear programs combined with classification algorithms. Our method allows for an ordinal level of compliance and can also be used to bound network treatment effects under interference and noncompliance. Critically, our approach does not require exclusion restriction or principal ignorability, ensuring its validity regardless of whether such assumptions hold. We reanalyze a cluster-randomized encouragement trial studying the effect of financial incentive for COVID-19 vaccine uptake in Ghana. Compliance with the financial incentive (i.e. encouragement) is measured using a three-level score of intention to take the vaccine. We detect effect heterogeneity across principal strata. Moreover, our sharp bounds are narrower than those using existing methods with covariate adjustment alone, or with linear programs and classification algorithms.

Instrumental Variables

Estimation of causal odds ratios in cluster randomized trials with non-compliance using structural nested models Xiangyu Yu* Xiangyu Yu, Nicholas Jewell,

We consider estimating the causal odds ratios using logistic structural nested mean models in cluster randomized trials with non-compliance and a dichotomous outcome. The randomization of assigning treatments provides an ideal instrumental variable for the unobserved confounding. Existing literature either limits the estimation using logistic structural nested mean models under the setting of single observations instead of clusters, or cannot yield consistent causally interpretable estimators due to the non-collapsibility of odds ratios. We propose both marginal and cluster-specific estimators that are guaranteed to estimate the causal odds ratios consistently under the null hypothesis, and circumvents computational challenges for implementation. We illustrate the proposed method in the context of a large-scale cluster randomized trial, Applying Wolbachia to Eliminate Dengue study, and perform the ‘as-treated’ analysis for both within-cluster and between-cluster effects.

Longitudinal Data Analysis

The causal effect of volatility: Estimation by marginal structural models Nanum Jeon*
Nanum Jeon, Ian Lundberg, Hao Liang,

U.S. workers operate in a labor market characterized by a high level of both inequality and volatility. Inequality exists to the degree that economic well-being is unequally distributed across people at a single point. Volatility exists to the degree that the economic well-being rises and falls for a given person over time. Both may be consequential, yet the causal effects of volatility have been understudied because volatility is a complicated causal treatment: it is by definition a trajectory experienced over time. We conceptualize the causal effect of volatility in the potential outcomes framework and show how to estimate using marginal structural models (MSMs). We develop new structural models designed to directly estimate the effects of volatility. We illustrate the method by studying the causal effect of employment volatility on marriage, and we close with a discussion of how our approach could unlock new research about the causal effects of volatility.

Machine Learning and Causal Inference

Hidden yet quantifiable: A lower bound for confounding strength using randomized trials
Piersilvio De Bartolomeis* Piersilvio De Bartolomeis, Javier Abad, Konstantin Donhauser, Fanny Yang,

In the era of fast-paced precision medicine, observational studies play a major role in properly evaluating new drugs in clinical practice. Yet, unobserved confounding can significantly compromise causal conclusions from observational data. We propose a novel strategy to quantify unobserved confounding by leveraging randomized trials. First, we design a statistical test to detect unobserved confounding with strength above a given threshold. Then, we use the test to estimate an asymptotically valid lower bound on the unobserved confounding strength. We evaluate the power and validity of our statistical test on several synthetic and semi-synthetic datasets. Further, we show how our lower bound can correctly identify the absence and presence of unobserved confounding in a real-world setting.

Machine Learning and Causal Inference

Strategic Decision-Making in the Presence of Information Asymmetry: Provably Efficient RL with Algorithmic Instruments Mengxin Yu* Mengxin Yu,

We study offline reinforcement learning under a novel model called strategic MDP, which characterizes the strategic interactions between a principal and a sequence of myopic agents with private types. Due to the bilevel structure and private types, strategic MDP involves information asymmetry between the principal and the agents. We focus on the offline RL problem, where the goal is to learn the optimal policy of the principal concerning a target population of agents based on a pre-collected dataset that consists of historical interactions. The unobserved private types confound such a dataset as they affect both the rewards and observations received by the principal. We propose a novel algorithm, Pessimistic policy Learning with Algorithmic iNstruments (PLAN), which leverages the ideas of instrumental variable regression and the pessimism principle to learn a near-optimal principal's policy in the context of general function approximation. Our algorithm is based on the critical observation that the principal's actions serve as valid instrumental variables. In particular, under a partial coverage assumption on the offline dataset, we prove that PLAN outputs a 1/sqrt{K}optimal policy with K being the number of collected trajectories. We further apply our framework to some special cases of strategic MDP, including strategic regression, strategic bandit, and noncompliance in recommendation systems.

Machine Learning and Causal Inference

A new parametrization of DAGs and causal Markov kernels for scientific feature discovery
Elise Walker* Elise Walker, Jonas Actor, Carianne Martinez, Nathaniel Trask,

Due to the complexity of multimodal scientific datasets, causal feature detection of these datasets necessitates unsupervised representation learning methods. Physics-informed multimodal autoencoders (PIMA) have recently demonstrated successful unsupervised feature detection with variational autoencoders (VAEs) where multiple scientific modalities and physics constraints acted as surrogates for the supervision labels typically needed for successful VAEs. Building upon the successes of PIMA, we present a VAE framework coupled with a trainable directed acyclic graph (DAG) to discover features with plausible causal relationships in multimodal scientific datasets. In particular, we introduce a new parametrization for learning both the edges of a DAG and the causal Markov kernels of the joint distribution of its nodes. We use this parametrization to simultaneously learn a DAG in conjunction with a latent space of a VAE. Training of our DAG and VAE is performed in an end-to-end differentiable framework via a single, tractable evidence lower bound (ELBO) loss function. We achieve a single ELBO by placing a Gaussian mixture prior on the latent space and identifying each of the Gaussians with an outcome of the DAG nodes. We demonstrate the efficacy of our DAG parametrization, and we test our joint VAE and DAG framework on both a synthetic and a scientific dataset. Our results demonstrate the capability of learning a DAG on discovered key features in an exploratory scientific setting.

Machine Learning and Causal Inference

Ultra-high dimensional confounder selection algorithms comparison with application to radiomics data Ismaila Balde* Ismaila Baldé, Debashis Ghosh,

Radiomics is an emerging area of medical imaging data analysis particularly for cancer. It involves the conversion of digital medical images into mineable ultra-high dimensional data. Machine learning algorithms are widely used in radiomics data analysis to develop powerful decision support model to improve precision in diagnosis, assessment of prognosis and prediction of therapy response. However, machine learning algorithms for causal inference have not been previously employed in radiomics analysis. In this paper, we evaluate the value of machine learning algorithms for causal inference in radiomics. We select three recent competitive variable selection algorithms for causal inference: outcome-adaptive lasso (OAL), generalized outcome-adaptive lasso (GOAL) and causal ball screening (CBS). We used a sure independence screening procedure to propose an extension of GOAL and OAL for ultra-high dimensional data, SIS + GOAL and SIS + OAL. We compared SIS + GOAL, SIS + OAL and CBS using simulation study and two radiomics datasets in cancer, osteosarcoma and gliosarcoma. The two radiomics studies and the simulation study identified SIS + GOAL as the optimal variable selection algorithm.

Machine Learning and Causal Inference

Performance of Cross-Validated Targeted Maximum Likelihood Estimation Matthew Smith*
Matthew Smith, Camille Maringe, Miguel Angel Luque Fernandez,

Background: Estimating causal relationships in public health is often of interest. Evidence shows that targeted maximum likelihood estimation (TMLE) often performs better than other estimators. However, TMLE suffers from variance underestimation due to overfitting the outcome model in the absence of the Donsker class condition. In such cases, cross-validated TMLE (CV-TMLE) is considered a suitable alternative to prevent overfitting and enhance variance estimation and could be beneficial in cases of near-positivity violations.

Methods: Using simulations, we compared different CV-TMLE strategies involving outcome model cross-validation or both outcome and exposure model cross-validation. We updated the user-friendly 'eltmle' Stata package with options for CV-TMLE, choice for the number of folds, retention of nuisance variables, and reporting of covariate balance tables.

Results and conclusion: Our results show CV-TMLE as a valid, and preferable, alternative to TMLE for variance estimation in the presence of near-positivity violations or data sparsity. Finally, we illustrate the benefits of CV-TMLE using an example from cancer epidemiology.

Machine Learning and Causal Inference

Strategic Decision-Making in the Presence of Information Asymmetry: Provably Efficient RL with Algorithmic Instruments Mengxin Yu* Mengxin Yu,

We study offline reinforcement learning under a novel model called strategic MDP, which characterizes the strategic interactions between a principal and a sequence of myopic agents with private types. Due to the bilevel structure and private types, strategic MDP involves information asymmetry between the principal and the agents. We focus on the offline RL problem, where the goal is to learn the optimal policy of the principal concerning a target population of agents based on a pre-collected dataset that consists of historical interactions. The unobserved private types confound such a dataset as they affect both the rewards and observations received by the principal. We propose a novel algorithm, Pessimistic policy Learning with Algorithmic iNstruments (PLAN), which leverages the ideas of instrumental variable regression and the pessimism principle to learn a near-optimal principal's policy in the context of general function approximation. Our algorithm is based on the critical observation that the principal's actions serve as valid instrumental variables. In particular, under a partial coverage assumption on the offline dataset, we prove that PLAN outputs a 1/sqrt{K}-optimal policy with K being the number of collected trajectories. We further apply our framework to some special cases of strategic MDP, including strategic regression, strategic bandit, and noncompliance in recommendation systems.

Machine Learning and Causal Inference

Dyadic Reinforcement Learning Shuangning Li* Shuangning Li,

Mobile health aims to enhance health outcomes by delivering interventions to individuals as they go about their daily life. The involvement of care partners and social support networks often proves crucial in helping individuals managing burdensome medical conditions. This presents opportunities in mobile health to design interventions that target the dyadic relationship — the relationship between a target person and their care partner — with the aim of enhancing social support. In this paper, we develop dyadic RL, an online reinforcement learning algorithm designed to personalize intervention delivery based on contextual factors and past responses of a target person and their care partner. Here, multiple sets of interventions impact the dyad across multiple time intervals. The developed dyadic RL is Bayesian and hierarchical. We formally introduce the problem setup, develop dyadic RL and establish a regret bound. We demonstrate dyadic RL's empirical performance through simulation studies on both toy scenarios and on a realistic test bed constructed from data collected in a mobile health study.

Machine Learning and Causal Inference

Missing Not At Random Data In Federated Learning Systems David Goetze* David Goetze, Rohit Bhattacharya, Jeannie Albrecht,

Federated learning is a technique used to train machine learning models on multiple datasets contained in local nodes without the need for information exchange between these nodes. Often these nodes correspond to multiple users operating on individual devices. A central server sends a model, e.g., an autocorrect model, to all devices, and intermittently pings them for model usage and accuracy. The user devices can choose to respond, sending back the loss or gradients allowing the central server to train and push an updated model back to all users. Thus federated learning enables data privacy while building a shared model across users.

Our focus is on federated learning for models trained using stochastic gradient descent, e.g., deep neural nets, when not all users choose to participate. Depending on why certain users do not share their data, this can lead to updates that cause the model to perform poorly across the general population (e.g., poor autocorrection for younger users if these users disproportionately choose not to participate). We examine missing data in federated learning through the lens of causal inference. In particular, we propose a reweighted version of stochastic gradient descent using the propensity score of missingness that unbiases the computation of gradients under Missing Not At Random assumptions. Finally, we present FLeWM, a federated learning system built to test our technique. We verify our results empirically on both simulated and real world data.

Machine Learning and Causal Inference

Proximal Causal Inference With Text Data Jacob M. Chen* Jacob M. Chen, Rohit Bhattacharya, Katherine A. Keith,

Recently, researchers have proposed causal inference methods that attempt to mitigate confounding bias by including unstructured text data as proxies of confounding variables that are partially or imperfectly measured. These approaches assume analysts have supervised labels of the confounders given text for at least a subset of instances, a constraint that is not always feasible due to data privacy or cost. In this work, we address settings in which an important confounding variable is completely unobserved. We propose a new causal inference method that splits pre-treatment text data, infers two proxies from two different zero-shot models on the separate splits, and plugs these proxies into the proximal g-formula. We prove that our particular method of text-based proxy generation satisfies the identification conditions required by the proximal g-formula while some other seemingly reasonable proposals do not. We evaluate our method in fully synthetic and semi-synthetic settings. For our semi-synthetic setting, we use real-world clinical notes from the MIMIC-III dataset and the Flan-T5 model, an instruction-tuned large language model, to infer proxies in a zero-shot manner. We find that our procedure results in causal estimates with low bias, whereas alternative procedures do not. This combination of proximal causal inference and zero-shot classifiers is novel to our knowledge and expands the set of text-specific causal methods available to practitioners.

Machine Learning and Causal Inference

From Paychecks to Plates: Tracing the Impact of Layoffs on Food Access in the United States Kiet Le* Kiet Le, Thanh Nguyen, Nina Rutledge,

The COVID-19 pandemic resulted in the numerous closures of businesses, caused widespread job losses, and negatively impacted the food security of US families. This study measures the extent to which food expenditure, food consumption, and the quality of food consumed by individuals were affected by layoffs. We estimated the treatment effect of involuntary job losses on these food security items using difference-in-difference regression, logistic regression, and augmented inverse propensity weighted estimation, and found that involuntary job losses caused Americans to be more vulnerable to food insecurity. Particularly, we found statistically significant evidence suggesting that layoffs caused people to eat less and switch to lower-quality food and unbalanced meals. The main purpose of this study is to shed light on the immediate economic implications of job loss on food security, which is having a ripple effect across public health, community development, and employment policies. We emphasize the importance of the social safety net and welfare programs that can provide immediate assistance to those who are directly impacted by layoffs.

Machine Learning and Causal Inference

Doubly Robust Estimation of Treatment Effect for Time-to-event Outcome under Dependent Left Truncation and Informative Right Censoring Yuyao Wang* Yuyao Wang, Andrew Ying, Ronghui Xu,

In aging studies or prevalent cohort studies, causal inference for time-to-event outcomes can be challenging. The challenges arise because, in addition to the potential confounding bias from observational data, the collected data usually also suffers from the bias by informative right censoring and the selection bias by left truncation, where only subjects with time to event (such as death) greater than the enrollment times are included. To assess the treatment effect on time-to-event outcomes in such settings, inverse probability weighting (IPW) is often employed. However, IPW is sensitive to model misspecifications, which makes it vulnerable, especially when faced with three sources of biases. Moreover, IPW is inefficient. To overcome these issues, we construct a model doubly robust estimator that have protection against the model misspecifications for all three sources of missing mechanisms, as well as a rate doubly robust estimator that is root-n consistent even when slower than root-n methods, such as nonparametric or machine learning methods, are incorporated. Our work represents the first attempt to construct doubly robust estimators that account for all three sources of biases: confounding bias, selection bias from covariate-induced dependent left truncation, and bias from informative censoring. We apply the proposed estimator to analyze the effect of midlife alcohol consumption on late life cognitive impairment using data from the Honolulu Asia Aging Study.

Machine Learning and Causal Inference

Distilling causal effects: stable subgroup estimation via distillation trees in causal inference Ana Kenney* Ana Kenney, Melody Huang, Tiffany Tang, Tanvi Shinkre,

We introduce a method, causal distillation trees (CDT), that allows researchers to stably estimate interpretable causal subgroups in their studies. CDT allows
researchers to fit any machine learning model of their choice to estimate the individual-level treatment effect, and then leverages a simple, second-stage tree-based model to then “distill’’ the estimated treatment effect into meaningful subgroups. As a result, we are able to leverage the theoretical guarantees from black-box machine learning models, while preserving the interpretability of a simple decision tree. We theoretically characterize the stability of CDT in estimating substantively meaningful subgroups, and provide helpful diagnostics for researchers to evaluate the quality of the estimated subgroups. We empirically demonstrate our method via extensive simulations and a case study on jobs training program experiments. We show that CDT out-performs state-of-the-art approaches in identifying interpretable subgroups.

Machine Learning and Causal Inference

Efficient, Cross-Fitting Estimation of Spatial Treatment Effects in Semiparametric Spatial Point Processes Xindi Lin* Xindi Lin, Hyunseung Kang,

Recently, there has been great interest in studying causal relationships in spatial settings where the observed data are non-i.i.d., spatial points on a grid. This paper studies the efficient estimation of treatment effects in a semiparametric spatial point process where the target estimand is the effect of a spatial covariate on the spatial distribution of points, and we allow for nonparametric adjustment of measured confounders. We generalize cross-fitting to spatial point processes. In particular, we use random thinning, a popular procedure in simulations and Bayesian MCMC, to split the spatial data and use spatial composite likelihoods for estimation. We show that our estimator is consistent and asymptotically Normal where the asymptotic variance can be consistently estimated. We also show that the proposed estimator achieves the semiparametric efficiency bound if the spatial point process is Poisson. We demonstrate the performance of our proposed method through a simulation study and a re-analysis of the spatial distribution of tree species. We show that compared to existing approaches based on parametric models, our approach provides a more robust, flexible, and, in some cases, efficient estimate of the target estimand.

Machine Learning and Causal Inference

DoubleLingo: Causal Estimation with Large Language Models Zach Wood-Doughty* Zach Wood-Doughty, Marko Veljanovski,

Estimating causal effects from non-randomized data requires assumptions about the underlying data-generating process. To achieve unbiased estimates of the causal effect of a treatment on an outcome, we must adjust for any confounding variables that influence both treatment and outcome. When such confounders include text data, existing causal inference methods struggle due to the high dimensionality of the text. The simple statistical models which have sufficient convergence criteria for causal estimation are not well-equipped to handle noisy unstructured text, but flexible Large language models (LLMs) that excel at predictive tasks with text data do not meet the statistical assumptions necessary for causal estimation. Our method enables theoretically consistent estimation of causal effects using LLM-based nuisance models by incorporating them within the framework of Double Machine Learning. On the best available dataset for evaluating such methods, we obtain a 10.4% reduction in the relative absolute error for the estimated causal effect over existing methods.

Machine Learning and Causal Inference

Causally Sufficient Dimension Reduction for Text Data Zach Wood-Doughty* Zach Wood-Doughty, Kayla Schroeder, Razieh Nabi,

Statistical modeling of text is complicated by the data's high dimensionality. Topic models are a central tool for representing text documents in low-dimensional space, and have been used in thousands of analyses of domains including literature and healthcare. While probabilistic topic models allow for human interpretation of large text corpora, the representations learned by these methods are inherently associational. Topic models reveal underlying structure within a corpus, but that structure does not necessarily represent the underlying causal structure that connects the text to other variables. When incorporating topic models as a piece of larger analyses, it is common to for researchers to manually interpret the topics and then fit models that link those topics to over variables of interest. However, if our goal is to understand causal relations between the text and other variables, we need to be explicit about the assumptions required to produce unbiased estimates of causal effects. This work applies causal sufficient dimensionality reduction to text data, enabling causal analyses of textual treatments.

Machine Learning and Causal Inference

Gaussian Processes for Social Scientists: A powerful tool for addressing model-dependency and uncertainty Soonhong Cho* Soonhong Cho, Doeun Kim, Chad Hazlett,

The Gaussian Process (GP) is a highly flexible but easy-to-understand tool for non-linear regression with rigorous handling of uncertainty estimation. Unlike a conventional parametric model, it accounts for uncertainty over model choice on the predicted values as we move away from the data, making it ideal for inference where poor overlap/model-dependency is an issue. GPs nevertheless remain underutilized in social science perhaps because (i) few resources for social scientists have explained them accessibly, and (ii) many existing software implementations of GPs are ill-suited to social science applications and require setting numerous hyperparameters. We begin by offering a simple but rigorous explanation for GPs rooted in a natural assumption—that observations that are closer in X will be closer in Y. We also provide a new implementation that improves interpretability and performance while avoiding most user-chosen hyperparameters. Next, while GPs are demonstrably not the best tool for all purposes, we describe and illustrate their advantages in contexts of high model-dependency/extrapolation, showing their more appropriate confidence intervals for conditional estimates (or at the limit, pointwise estimates). We also illustrate their performance by simulation and empirical studies in (i) imputation-based treatment effect estimation where parametric models perform poorly, and (ii) regression discontinuity designs (RDD) under different causal assumptions.

Machine Learning and Causal Inference

Efficient generalizability and transportability of survival causal effects Axel Martin* Axel Martin, Ivan Diaz, Michele Santacatterina,

Randomized clinical trials (RCTs) are widely regarded as the gold standard for estimating survival causal effects. However, the subpopulation included in RCTs rarely reflects real-world populations. As a result, the generalization or transportability of results from RCTs to those populations of interest is limited. Real-world observational data, on the other hand, are generally widely available and contain a substantial amount of information regarding the population of interest.
Recent works have proposed leveraging observational data to generalize and transport causal effects. However, few of these methods focus on survival causal effects, have been widely understudied in the literature. We contribute to this important literature by proposing a flexible and robust doubly-robust estimator that incorporates machine learning techniques to transport and generalize survival causal effects from RCTs to real-world target populations. Additionally, we propose more efficient estimators by leveraging treatment effect modification. We demonstrate their large sample properties, evaluate their finite sample performance in simulations, and apply them using the Women’s Health Initiative RCT and observational studies. Finally, we have developed an open-source R package, ‘dmlSurv,’ to implement these estimators, facilitating their accessibility and utilization.

Machine Learning and Causal Inference

Integrating Detection and Causal Models using Fully Latent Principal Stratification Kirk Vanacore* Kirk Vanacore, Adam Sales,

Predictive models can provide real-time estimations of human attributes, dispositions, propensities, or attitudes. When implemented in computer applications, they can provide insight into users' latent states. However, they do not address the optimal actions these systems should take given the predictions. Fully Latent Principal Stratification (FLPS) provides one solution by allowing for the estimation of subgroup effects when those subgroups are determined after randomization and defined by a latent variable. The integration of detection models and FLPS have the potential to turn predictions into actionable recommendations based on estimated causal effects.

We illustrate this application using an example study of a detection model implementation in a Computer-Based Learning Platform (CBLP). In this example we address the issue of 'gaming the system' – a behavior categorized by attempting to progress through a learning activity without learning – in a CBLPs. Using the combination of a 'gaming the system' detection model and FLPS, we are able to estimate how students who have a high propensity to game the system in one CBLP would respond to modifications of the original CBLP as well as CBLPs with different pedagogical approaches. Through this analysis, we determine that simple manipulations of feedback systems within a CBLP may be more optimal than global changes, such as gamification.

Machine Learning and Causal Inference

C-Learner: Constrained Learning for Causal Inference Tianhui (Tiffany) Cai* Tianhui (Tiffany) Cai, Yuri Fonseca, Kaiwen Hou, Hongseok Namkoong,

A fundamental problem in causal inference is the accurate estimation of the average treatment effect (ATE). Existing methods such as Augmented Inverse Probability Weighting (AIPW) and Targeted Maximum Likelihood Estimation (TMLE) are asymptotically optimal. Although these methods are asymptotically equivalent, they exhibit significant differences in finite-sample performance, numerical stability, and complexity, which raises questions about their relative practical utility.

In response, we develop the Constrained Learner (C-Learner), which is a new asymptotically optimal method for estimating the ATE. C-Learner is flexible and conceptually very simple: it directly encodes the condition for asymptotic optimality of the estimator as a constraint for learning outcome models, which are then used in a plug-in estimator for the ATE. C-Learner can thus leverage tools and advances from constrained optimization to learn these outcome models. In practice, we find that C-Learner performs comparably to or better than other asymptotically optimal methods. These attributes collectively position C-Learner as a compelling new tool for researchers and practitioners of causal inference.

Machine Learning and Causal Inference

A tutorial for studying continuous, time-varying, and/or otherwise complex exposures using longitudinal modified treatment policies Katherine Hoffman* Katherine Hoffman, Ivan Diaz, Nicholas Williams, Kara Rudolph, Diego Salazar-Barreto,

This tutorial discusses methodology for causal inference using longitudinal modified treatment policies (LMTPs). LMTPs facilitate the mathematical formalization, identification, and estimation of many novel parameters, and mathematically generalize many commonly used parameters, such as the average treatment effect. LMTPs apply to a wide variety of exposures, including binary, multivariate, and continuous, and can accommodate time-varying treatments and confounders, competing risks, loss-to-follow-up, as well as survival, binary, or continuous outcomes. LMTPs can be seen as an extension of static and dynamic interventions to involve the natural value of treatment, and, like dynamic interventions, can be used to define alternative estimands with a positivity assumption that is more likely to be satisfied than estimands corresponding to static interventions. This tutorial aims to illustrate several practical uses of the LMTP methodology, including describing different estimation strategies and their corresponding advantages and disadvantages. We provide numerous examples of types of research questions which can be answered using LMTPs. We go into more depth with one of these examples—specifically, estimating the effect of delaying intubation on critically ill COVID-19 patients' mortality. We demonstrate the use of the open-source R package {lmtp} to estimate the effects, and we provide code on https://github.com/kathoffman/lmtp-tutorial.

Machine Learning and Causal Inference

HAL-based plugin estimation of the Causal Dose–Response Curve Junming (Seraphina) Shi*
Junming (Seraphina) Shi, Alan Hubbard, Mark Van Der Laan,

Estimating the marginally adjusted dose–response curve for continuous treatments is a longstanding statistical challenge critical across multiple fields. In the context of parametric models, misspecification may result in substantial bias, hindering the accurate discernment of the true data generating distribution and the associated dose-response curve. In contrast, non-parametric models face difficulties as the dose-response curve isn't pathwise differentiable, and then there is no √n-consistent estimator. The emergence of the Highly Adaptive Lasso (HAL) MLE by van der Laan[2015] and van der Laan [2017] and the subsequent theoretical evidence by van der Laan [2023] regarding its pointwise asymptotic normality and uniform convergence rates, have highlighted the asymptotic efficacy of the HAL-based plug-in estimator for this intricate problem. This research delves into the HAL-based plug-in estimators, including those with cross-validation and undersmoothing selectors, and introduces the undersmoothed smoothness-adaptive HAL-based plug-in estimator. We assess these estimators through extensive simulations, employing detailed evaluation metrics. Building upon the theoretical proofs in van der Laan [2023], our empirical findings underscore the asymptotic effectiveness of the undersmoothed smoothness-adaptive HAL-based plug-in estimator in estimating the marginally adjusted dose–response curve.

Machine Learning and Causal Inference

Two-Step Targeted Minimum-Loss Based Estimation for Non-Negative Two-Part Outcomes
Nicholas Williams* Nicholas Williams, Kara Rudolph, Ivan Diaz,

Non-negative two-part outcomes are defined as outcomes with a density function that have a zero point mass but are otherwise positive. Examples, such as healthcare expenditure and hospital length of stay, are common in healthcare utilization research. Despite the practical relevance of non-negative two-part outcomes, very few methods exist to leverage knowledge of their semicontinuity to achieve improved performance in estimating causal effects. In this paper, we develop a nonparametric two-step targeted minimum-loss based estimator (denoted as hTMLE) for non-negative two-part outcomes. We present methods for a general class of interventions referred to as modified treatment policies, which can accommodate continuous, categorical, and binary exposures. The two-step TMLE uses a targeted estimate of the intensity component of the outcome to produce a targeted estimate of the binary component of the outcome that may improve finite sample efficiency. We demonstrate the efficiency gains achieved by the two-step TMLE with simulated examples and then apply it to a cohort of Medicaid beneficiaries to estimate the effect of chronic pain and physical disability on days' supply of opioids.

Machine Learning and Causal Inference

Adapting causal forests for practical challenges: an application to randomised controlled trials Eleanor Van Vogt* Eleanor Van Vogt, Suzie Cro, Karla Diaz-Ordaz,

Randomised controlled trials (RCTs) typically focus on estimating the average treatment effect (ATE), often resulting in null conclusions. Where treatment heterogeneity is of interest, the variables for which heterogeneity is suspected need to be pre-specified, to avoid “data dredging”. Nevertheless, there is an increased interest in exploring personalised treatments and policy decisions, often done through estimating heterogeneous treatment effects (HTEs).
Causal machine learning methods for HTEs have grown in popularity, offering flexible tools for exploring heterogeneity in complex settings using machine learning to learn the conditional average treatment effect (CATE).
Here, we use (Bayesian) causal forests to analyse two completed RCTs to explore potential treatment effect heterogeneity. We explored adaptations needed for the methods to work in challenging applied settings, such as rare outcomes and missing data. To identify simple data-driven subgroups delineating mal-responders, we proposed a method based on median root split, contrasted with dichotomising continuous variables from the beginning. We also tested permutation-based variable importance.
We discuss potential modifications to handle competing risks, such as the incorporation of subdistribution weights and imputation approaches. Additionally, we highlight the importance of cross-fitting and sample splitting concerning accurate quantification of standard error in generating HTE results.

Machine Learning and Causal Inference

Off-Policy Learning of Content Promotions: Optimizing Digital Distribution Channels Joel Persson* Joel Persson,

A common decision-making problem for digital content publishers is deciding which content to promote on their distribution channels (e.g., digital front page, email newsletters, and social media pages), where a small set of items is shown to the entire user base and learning by randomly promoting content is typically undesirable from a business perspective. In this paper, we develop an off-policy learning framework for this decision problem. Our objective is to learn an optimal policy per channel, which, given information about content, decides which items to promote in order to maximize the mean performance among all content. Building on methods from causal machine learning, we show that counterfactual policies are non-parametrically identified from historical data and that the optimal policy selects the content with the largest conditional average treatment effect of promotion. We then present a data-driven estimation procedure that is scalable and doubly robust. To evaluate our framework, we partnered with an international newspaper, where the goal is to optimize the selection of content to promote on its digital front pages, and show that our optimal policy outperforms common baselines and significantly improves over the company’s current practice. Altogether, our work provides a framework for offline evaluation of data-driven promotion strategies for digital content providers and thereby supports optimizing their distribution channels.

Matching

Treatment bootstrapping: a new approach to quantify uncertainty of average treatment effect estimates Jing Li* Jing Li,

This paper proposes a new non-parametric bootstrap method to quantify the uncertainty of average treatment effect estimates from matching estimators. More specifically, it seeks to quantify the uncertainty associated with the average treatment effect estimate for the treated by bootstrapping the treatment group only and finding the counterpart control group by matching on estimated propensity score. We demonstrate the validity of this approach and compare it with existing bootstrap approaches through Monte Carlo simulation and real world example data. The results indicate that the proposed approach constructs confidence intervals that have comparable precision and coverage rate as existing bootstrap approaches and can produce smaller standard error estimates albeit with lower coverage rate depending on the proportion of treatment group units in the sample data and the specific matching method used.

Matching

Flexible Almost-Exact Matching for Trustworthy Causal Inference Quinn Lanners* Quinn Lanners, Harsh Parikh, Alexander Volfovsky, Cynthia Rudin, David Page, Brandon Westover, Sahar Zafar, Zade Akras,

Applying observational causal inference methods to real-world problems is difficult for a variety of reasons. We are often working with noisy and messy datasets and are forced to rely on untestable assumptions. Given this reality, our goal is to produce methods for observational causal inference that are auditable, easy to troubleshoot, yield accurate treatment effect estimates, and are adaptable to various datasets. We describe an almost-exact matching approach that achieves these goals by (i) learning a distance metric via outcome modeling, (ii) creating matched groups using the distance metric, and (iii) using the matched groups to estimate treatment effects. Our proposed method uses variable importance measurements to construct a distance metric, making it a flexible method that can be adapted to various applications. We operationalize this method into a safe and interpretable framework to identify optimal treatment regimes in a noisy ICU dataset. In this application, we face challenges including missing data, inherent stochasticity, and the critical requirements for interpretability and patient safety. Using our approach, we match patients with similar medical and pharmacological characteristics, allowing us to construct an optimal policy via interpolation. Our findings strongly support personalized treatment strategies based on a patient’s medical history and pharmacological features.

Matching

Caliper Synthetic Matching: Generalized Radius Matching with Local Synthetic Controls
Xiang Meng* Xiang Meng, Luke Miratrix, Jonathan Che,

Matching promises transparent causal inferences for observational data, making it an intuitive approach for many applications. In practice, however, standard matching methods often perform poorly compared to modern approaches such as response-surface modeling and balancing. Specifically, when potential outcomes are non-linear functions of covariates, and covariate distribution for treatments and controls are balanced marginally but not jointly, methods that only guarantee marginal balance may produce biased estimates. We propose Caliper Synthetic Matching (CSM) to address these challenges while providing provable bounds for joint balance, preserving simple and transparent matches and match diagnostics. CSM improves the existing matching methods by incorporating a more general distance metrics, adaptive calipers, and locally constructed synthetic control weights. We show that CSM can be viewed a member in the monotonic imbalance bounding (MIB; Iacus et al., 2011) class , with a consequence of controlling the joint covariate balance. Compared to Coarsened Exact Matching (CEM; Iacus et al., 2012), another member in the MIB class, CSM offers an improved bias bound due to tighter local matching and synthetic control weights. Using a simulation study, we illustrate how CSM can outperform modern matching methods in settings mentioned earlier. Finally, we illustrate its usage on canonical datasets and provide guidelines in practice.

Measurement error and missing data

Measurement Error in Causal Inference: A Review Kevin Josey* Keith Barnatchez, Kevin Josey, Rachel Nethery,

In both the scientific application and development of causal inference methods, it is often implicitly assumed that all relevant variables are measured without error. Despite the extensive literature studying the impact of measurement error in association studies, the development of methods at the intersection of measurement error and causal inference is in a relatively early, yet rapidly growing, stage. In this paper, we provide an overview of the burgeoning field of measurement error in causal inference. We detail the key role of study design in addressing measurement error, before examining a variety of methods for addressing confounder and exposure measurement error in causal inference studies, synthesizing the existing methods in measurement error correction with common causal estimators of the means of the potential outcomes. To facilitate the comparison of existing methods and development of future methods, we frame all methods in terms of causal assumptions and their associated study design requirements. After introducing the different choices for measurement error correction, we conduct a simulation study to evaluate their relative merits. We conclude with a set of recommendations for causal inference researchers suspecting measurement error in their analysis.

Mediation

Identifying correlate of protection for COVID-19 vaccines using causal inference framework Nancy Hiu Lan Leung* Nancy Hiu Lan Leung,

The COVID-19 pandemic has resulted in significant loss of life, and vaccination has been instrumental in helping us return to a new normal. Immune biomarkers, for example neutralizing antibodies, are used to predict COVID-19 vaccine efficacy and to evaluate updated vaccines in immunobridging studies, i.e. these immune biomarkers are considered as correlate of protection (CoP). However, in a recent randomized trial of third-dose COVID-19 vaccination, we found that neutralizing antibodies did not associate with vaccine efficacy for inactivated COVID-19 vaccines. A significant hurdle to identify new CoP for inactivated vaccines is the difficulty to attribute causal effect of individual CoP to protection; but so far only limited research has been conducted to address this analytic challenge. We are conducting additional serologic testing of different immune biomarkers against circulating viruses in infected and uninfected individuals from our several on-going longitudinal cohorts and vaccine trials of COVID-19 vaccination in Hong Kong, and we also collect other relevant data such as demographics and prior vaccination and infection history. We will evaluate causal diagrams that reflect different hypotheses of immune protection with these data. We will evaluate the potential role of individual immune biomarkers as CoP by estimating the causal effect of vaccination on protection mediated by each biomarker.

Mediation

Causal Mediation Analysis with Hidden Mediators Laura Montelisciani* Laura Montelisciani, Eric Tchetgen Tchetgen,

Introduction: In causal mediation analysis, assumptions regarding the presence of a measurable mediator without measurement error are often impractical with observational data. A recent proximal causal inference framework enables the estimation of natural direct (NDE) and indirect (NIE) effects of the treatment on the outcome, even with measurement error or a hidden mediator.

Estimation: Two mediator proxies, Z and W, meeting specific conditional independence conditions, are required. Z and W must be directly caused by the mediator and associated with the outcome only via the mediator. For a continuous outcome, two linear models are fitted to estimate the NDE. The first determines the conditional expectation of W on treatment and Z, and the second, derives an unbiased estimator for the NDE from the regression of the outcome on the plugged in predicted conditional expectation of W and the treatment.
The estimate of NIE is then obtained by subtracting the estimate of NDE from the Average Treatment Effect, where the latter is estimated using the observed data. Estimation of NDE and NIE is possible even in the presence of interaction between the treatment and the hidden mediator.

Discussion: This estimation approach offers the benefit of accurately estimating both NDE and NIE in situations involving a hidden/mismeasured mediator, utilizing straightforward linear models and avoiding assumptions on measurement error.

Mediation

A separable effects approach to identifying the impact of prenatal anesthesia exposure on childhood behavior disorders Amy Pitts* Caleb Miles, Amy Pitts, Ling Guo, Caleb Ing,

The U.S. Food and Drug Administration has cautioned that prenatal exposure to anesthetic drugs during the third trimester may have neurotoxic effects; however, there is limited clinical evidence available to substantiate this recommendation. To explore this claim, we analyze data from the nationwide Medicaid Analytic eXtract (MAX) from 1999 through 2013, which linked 16,778,281 deliveries to mothers enrolled in Medicaid during pregnancy. The goal of our analysis is to estimate the causal effect of exposure to anesthesia in utero on the diagnosis of attention-deficit/hyperactivity disorder (ADHD) in the child. Isolating the effect of anesthesia from the effect of the surgical procedure is challenging due to these exposures being deterministically linked, thereby inducing an extreme positivity violation. To overcome this, we adopt the separable effects framework of Robins and Richardson (2010) to isolate the effect of anesthesia by blocking effects through variables that are assumed to completely mediate the causal pathway from surgery to DIBD. Furthermore, we develop sensitivity analyses to assess the impact of violations to our key identifying assumptions.

Mediation

Statistical Inference for High Dimensional Mediation Effects Yuzhou Lin* Yuzhou Lin, Xihong Lin,

Evaluating the effect of a treatment on an outcome via a mediator has received growing attention in clinical and genetic studies. Traditional mediation effect testing methods, including the Wald-type Sobel's test and the Joint Significance test, suffer from overconservative type-I-error and low power under a great quantity of composite null hypotheses. The recently developed divide-aggregate-composite-null test (DACT) properly controls the type-I-error with high power when any of its composite null case has proportion close to 1. But DACT's performance in other settings is unclear. We showed that under unfavorable settings, when no case has proportion close to 0 or when the effect size is large, DACT will fail to control the type-I-error, even with its default normal calibration under Efron's empirical null framework. We proposed a new calibration involving a three-component mixture model for DACT. We controlled the type-I-error while preserving high power compared with state-of-the-art testing methods under both favorable and unfavorable settings. A new procedure for estimating null proportions and a variation of DACT is proposed to boost its null estimation accuracy and power.

Mediation

Separable pathway effects for semi-competing risks in multi-state models, with application to leukemia data Yuhao Deng* Yuhao Deng, Yi Wang, Xiao-Hua Zhou,

Semi-competing risks refer to the phenomenon where a primary event (such as mortality) can "censor" an intermediate event (such as relapse of a disease), but not vice versa. Under the multi-state model, the primary event is decomposed to a direct outcome event and an indirect outcome event through intermediate events. Within this framework, we show that the total treatment effect on the cumulative incidence of the primary event can be decomposed into three separable pathway effects, corresponding to treatment effects on population-level transition rates between states. We next propose estimators for the counterfactual cumulative incidences of the primary event under hypothetical treatments by generalized Nelson-Aalen estimators with inverse probability weighting, and then derive the asymptotic normality of these estimators. Finally, we propose hypothesis testing procedures on these separable pathway effects based on logrank statistics. As an illustration of its potential usefulness, the proposed method is applied to compare effects of different allogeneic stem cell transplantation types on overall survival after transplantation. We find that haploidentical transplantation significantly reduces the risk of mortality by reducing the risk of relapse.

missing data

Zero Inflation as a Missing Data Problem: a Proxy-based Approach Trung Phung* Trung Phung, Jaron J.R. Lee, Opeyemi Oladapo-Shittu, Eili Y. Klein, Ayse P. Gurses, Susan M. Hannum, Kimberly Weems, Jill Marsteller, Sara E. Cosgrove, Sara C. Keller, Ilya Shpitser,

Zero-inflated data has values incorrectly recorded as zeros due to data recording conventions (rare outcomes assumed to be absent) or details of data recording equipment (artificial zeros in genomic data).
Common statistical models for zero-inflated data are parametric and generally assume at most missing-at-random. On the other hand, graphical missing data models are nonparametric and may handle missing-not-at-random, yet they require censored values to be marked by a special symbol like “?”, while “0” denotes both true and missing value in zero-inflated data.
This paper views zero-inflated data as a harder type of missing data, where a missingness indicator is unobserved whenever a zero is recorded. We show that in most cases, target parameters involving a zero-inflated variable are nonparametrically unidentified. However, if a proxy of the censoring indicator is observed, a modification of the Kuroki and Pearl’s effect restoration allows identification and estimation, given the proxy-indicator relationship is known.
If this relationship is unknown, our approach yields a partial identification for sensitivity analysis. Specifically, only certain proxy-indicator conditionals are compatible with the observed data distribution. We give an analytical bound for binary cases, while for more complex cases, Duarte et (2023)’s numerical bound should be computed.
We illustrate our method via simulation studies and a data application on Central Line-Associated Bloodstream Infections.

Observational studies

Confounder Selection via Iterative Graph Expansion Richard Guo* Richard Guo, Qingyuan Zhao,

Confounder selection, namely choosing a set of covariates to control for confounding between a treatment and an outcome, is arguably the most important step in the design of observational studies. Previous methods, such as Pearl's celebrated back-door criterion, typically require pre-specifying a causal graph, which can often be difficult in practice. We propose an interactive procedure for confounder selection that does not require pre-specifying the graph or the set of observed variables. This procedure iteratively expands the causal graph by finding what we call "primary adjustment sets" for a pair of possibly confounded variables. This can be viewed as inverting a sequence of latent projections of the underlying causal graph. Structural information in the form of primary adjustment sets is elicited from the user, bit by bit, until either a set of covariates are found to control for confounding or it can be determined that no such set exists. Other information, such as the causal relations between confounders, is not required by the procedure. We show that if the user correctly specifies the primary adjustment sets in every step, our procedure is both sound and complete. A Shiny web-app is available for practitioners to select confounders in an interactive fashion: https://ricguo.shinyapps.io/InteractiveConfSel/

Optimal Dynamic Treatment Effects

Inference on Optimal Dynamic Policies via Softmax Approximation Vasilis Syrgkanis* Qizhao Chen, Morgane Austern, Vasilis Syrgkanis,

Estimating optimal dynamic policies from offline data is a fundamental problem in dynamic decision making. In the context of causal inference, the problem is known as estimating the optimal dynamic treatment regime. Even though there exists a plethora of methods for estimation, constructing confidence intervals for the value of the optimal regime and structural parameters associated with it is inherently harder, as it involves non-linear and non-differentiable functionals of unknown quantities that need to be estimated. Prior work resorted to sub-sample approaches that can deteriorate the quality of the estimate. We show that a simple soft-max approximation to the optimal treatment regime, for an appropriately fast growing temperature parameter, can achieve valid inference on the truly optimal regime. We illustrate our result for a two-period optimal dynamic regime, though our approach should directly extend to the finite horizon case. Our work combines techniques from semi-parametric inference and g-estimation, together with an appropriate triangular array central limit theorem, as well as a novel analysis of the asymptotic influence and asymptotic bias of softmax approximations.

Partial Identification

Partial Identification of Causal Effects Using Proxy Variables AmirEmad Ghassami* AmirEmad Ghassami, Ilya Shpitser, Eric Tchetgen Tchetgen,

Proximal causal inference is a recently proposed framework for evaluating causal effects in the presence of unmeasured confounding. For point identification of causal effects, it requires identification of certain nuisance functions called bridge functions using proxy variables that are sufficiently relevant to the unmeasured confounder, formalized as a completeness condition. However, completeness is not testable, and although a bridge function may exist, lack of completeness may severely limit prospects for identification of a bridge function and thus a causal effect; therefore, restricting the application of the framework. In this work, we propose partial identification methods that do not require completeness and obviate the need for identification of a bridge function, i.e., we establish that proxies can be leveraged to obtain bounds on the causal effect even if available information does not suffice to identify a bridge function. Our bounds are non-smooth functionals of the underlying distribution. Hence, in the context of inference, we initially employ the LogSumExp approximation to obtain smooth approximations of our bounds. Subsequently, we leverage bootstrap confidence intervals on the approximated bounds. We further establish analogous results in related settings where identification hinges upon hidden mediators for which proxies are available, yet such proxies are not sufficiently rich for point identification of a bridge function or a corresponding causal effect.

Proximal causal inference

Proximal causal inference for continuous point-exposure treatments Antonio Olivas-Martinez* Antonio Olivas-Martinez, Andrea Rotnitzky,

The recently introduced proximal causal inference framework has revolutionized the identification and doubly-robust inference of causal effects, particularly in scenarios where confounding arises from an unobserved variable, yet observed proxy variables for the confounder are available. In such scenarios, estimating the parameter of interest requires estimating nuisance functions which are solutions to integral equations. While various machine learning methods exist for estimating these nuisance functions for binary treatments, limited attention has been given to the case of continuous point-exposure treatments. In this work, we extend the proximal causal inference framework to encompass continuous point-exposure treatments, employing a minimax optimization approach for estimating the nuisance functions. Our work fills a critical gap in the literature by demonstrating how finite sample properties of the estimator vary with the level of correlation between the proxy and unobserved confounders. To illustrate the method’s applicability, we apply it to the context of COVID-19 vaccines, focusing on identifying correlates of protection for use in immunobridging trials.

Randomized Studies

Comparing HIV Vaccine Immunogenicity across Trials with Different Populations and Study Designs Yutong Jin* Yutong Jin, Alex Luedtke, David Benkeser,

Recent studies has revealed that effective vaccines contributed tremendously to the prevention of infectious diseases. The effectiveness of vaccine is typically measured in randomized efficacy trials using clinical endpoints. Due to the difficulty of collecting outcomes on everybody within a large trial, this process is often time-consuming and has significantly slowed the vaccine research. To reduce the development time, one promising solution is to identify and assess surrogate endpoints, like immune responses, that are predictive of vaccine efficacy. However, the measurement of such immune responses is expensive so that a two-stage sampling strategy is commonly employed for large trials. Additionally, different clinical trials are conducted in diverse study populations across the world. It is thus hard to provide an objective comparison of vaccine immunogenicity between different HIV regimens directly. To address these challenges, we propose a framework that is capable of identifying appropriate causal estimands and estimators, which can be used to provide standardized comparisons of vaccine immunogenicity across trials. We evaluate the performance of the proposed estimands via extensive simulation studies. Our estimators are well-behaved and enjoy robustness properties. The proposed technique is applied to data from several HIV vaccine trials.

Randomized Studies

Do algorithms help humans make better decisions? A framework for experimental evaluation Sooahn Shin* Melody Huang, Eli Ben-Michael, D. James Greiner, Kosuke Imai, Zhichao Jiang, Sooahn Shin,

The use of data-driven algorithms has become ubiquitous in today's society. And yet, in many cases, humans still make final decisions especially when stakes are high. The critical question, therefore, is whether or not algorithms help humans make better decisions. We introduce a new methodological framework that can be used to experimentally evaluate the causal impact of algorithmic recommendations on human decisions. We first formalize decision maker's ability to make correct decisions using standard classification metrics based on potential outcomes. We then consider an experiment, where cases are randomly assigned to human decisions, either with or without algorithmic recommendations. We show how to analyze the data from such an experiment to compare the performance of three alternative decision-making systems–human alone, human with algorithm, and algorithm alone–under a minimal set of assumptions. We develop a sensitivity analysis to assess the degree to which the empirical estimates are robust to potential violations of the underlying assumptions. We apply the proposed methodology to the first randomized control trial of pretrial public safety assessment and examine whether or not the provision of risk assessment scores improve a judge's decision to impose a cash bail.

Randomized Studies

Causal Inference for Balanced Incomplete Block Designs Taehyeon Koo* Taehyeon Koo, Nicole Pashley,

Researchers often turn to block randomization to increase the precision of their inference or due to practical considerations, such as in multi-site trials. However, if the number of treatments under consideration is large it might not be practical or even feasible to assign all treatments within each block. We develop novel inference results under the finite-population design-based framework for a natural alternative to the complete block design that does not require reducing the number of treatment arms, the Balanced Incomplete Block Design (BIBD). This includes deriving the properties of two estimators under BIBDs and proposing conservative variance estimators. To assist practitioners in understanding the trade-offs of using BIBDs over other designs, the precisions of resulting estimators are compared to standard estimators for the complete block design and the completely randomized design. Simulations and data illustration demonstrate the strengths and weaknesses of using BIBDs. This work highlights BIBDs as practical and currently underutilized designs.

Randomized Studies

Rothman sufficient cause urn analysis of the truncation by death problem Jaffer Zaidi* Jaffer Zaidi,

The analysis of causal effects when the outcome of interest is possibly truncated by death has a long history in statistics. The survivor average causal effect is commonly identified with more assumptions than those guaranteed by the design of a randomized clinical trial. This paper demonstrates that individual level causal effects in the `always survivor' principal stratum can be identified with no stronger identification assumptions than randomization. Our methods are applied to randomized clinical trials in oncology, HIV maternal-to child transmission, and severe Twin-to-Twin transfusion syndrome.

Randomized Studies

Efficient Design-based Inference for Stepped Wedge Designs Fan Xia* Fan Xia, James Hughes, Patrick Heagerty, Gary Chan, Emily Voldal, Avi Kenny,

Stepped wedge designs (SWD) are a type of cluster randomized trial used to evaluate new interventions on clusters like clinics and communities. In typical SWDs, all clusters start in the control group, crossover to the intervention group at randomized times, and remain until the trial ends. The unique trial structure introduces challenges due to temporal confounding and opportunities through randomized crossovers, demanding a balanced approach to optimize data use for both robustness and efficiency. Linear mixed models (LMMs), common in SWDs, require restrictive modeling assumptions. Nonparametric methods are less assumption-heavy but often lack efficiency. Following the intuition that vertical comparisons within time points can help establish robustness, and horizontal comparisons between time points can increase power, we propose an estimator based on the semiparametric efficiency theory for clustered data. The proposed estimator is robust to misspecification in the outcome model (e.g. mis-specified temporal trend) and the covariance matrix. Moreover, it is semiparametrically efficient when both are correctly specified. For inference, we propose both a sandwich-type variance estimator and a conservative plug-in estimator from exact variance enhanced with a leave-one-out correction for finite sample bias, given the typically small number of clusters in SWDs. Additionally, an unbiased estimator is introduced to further correct bias in the conservative estimator.

Randomized Studies

Weighting-Based Estimators for the Survivor Average Causal Effect in Cluster Randomized Trials Dane Isenberg* Dane Isenberg, Fan Li, Nandita Mitra, Michael Harhay,

Studies often examine the effect of a binary treatment on a non-mortal outcome, where participants may have truncated outcomes if they do not survive to follow-up. One approach to address truncation by death is to target the survivor average causal effect (SACE), a casually interpretable and estimable conditional treatment effect defined via principal stratification. However, current uses for SACE either focus on individual randomized trials or require strong distribution assumptions when applied to multilevel data. We develop a SACE framework for cluster randomized trials (CRTs) relaxing restrictions on the distributions. We establish sets of assumptions that address latent confounding due to clustering to enable point identification of SACE for CRTs. We propose weighting-based estimators for SACE and provide asymptotic variance expressions when survival status is modeled using a GLMM. In simulations, we evaluate our estimators demonstrating that they account for latent confounding and are robust to certain departures from assumptions. We apply our methods to a CRT assessing the impact of a sedation protocol on mechanical ventilation among children with acute respiratory failure.

Randomized Studies

Asymptotically Valid Permutation Test in Strongly Mixing Conditions David Kim* David Kim, EunYi Chung,

Two-sample permutation test is might be used on comparing distribution of potential outcomes under binary treatments. However, permutation tests usually considered only on independent samples, which might be unrealistic assumption in Econometrics. Moreover, permutation tests might be not asymptotically valid when we consider the weaker null on parameters. Here we introduce a novel, asymptotically valid two-sample permutation test on parameters θ of P and Q, where each samples from distribution need not be independent – they only assumed to be strongly mixing with other mild moment conditions. We do this by implementing both recent functional central limit theorem on non-stationary process and self-normalization on the test statistic. Particularly, this can be extended to the test for causal estimands like ATE in dependent structure. We give a Monte Carlo simulation study to check and compare the finite sample performance of this new test with standard normal approximation, fixed-b asymptotics, and moving block bootstrap. We further give simulations on some block-wise permutation test, which we might discover its nature in the future.

Randomized Studies

ML-assisted Randomization Tests for Complex Treatment Effects in A/B Experiments
JungHo Lee* JungHo Lee, Wenxuan Guo, Panos Toulis,

In recent years, online businesses have utilized large-scale experimentation for data-driven decision making. In an A/B experiment, for example, a business randomizes two different treatments (e.g., website designs) to their customers and then aims to infer which treatment is better. However, most existing approaches, including those that leverage flexible machine learning (ML) tools, are valid only in an asymptotic regime that does not exploit the controlled randomness in the experiment. As a result, treatment effect estimation with these methods can be biased, or even impossible in settings where the experimental design is complex. In this paper, we develop randomization tests for several treatment effects of interest, including complex effects such as heterogeneity and interference. These tests exploit the true experimental variation and are thus guaranteed to always be finite-sample valid. In addition, we construct our randomization tests using flexible ML models, where the test statistic is defined as the difference between the cross-validation errors from two ML models, one with and another without the complex treatment effect. This construction effectively combines the finite-sample validity of the randomization framework with the prediction power of modern ML tools. We demonstrate this combined benefit via extensive synthetic and real-world experiments.

Regression Discontinuity

Blessing of Multiple Control Groups in Fuzzy Regression Discontinuity Designs: Evaluating Extended Time Accommodations Youmi Suk* Youmi Suk, Yongnam Kim,

Regression discontinuity (RD) designs have become popular for evaluating the effectiveness of policies and programs, and in particular, fuzzy RD designs are often employed in the presence of noncompliance. While recent advancements have combined RD designs with other quasi-experimental designs, there is limited research on enhancing RD designs with multiple control groups. This paper proposes a novel approach to combining fuzzy RD designs and multiple control group designs under one-sided noncompliance. We estimate the average treatment effect on the treated (ATT) at the cutoff from the fuzzy RD design under one-sided noncompliance and construct its bounds derived from multiple control groups. Using the bounds as a sensitivity check, we examine whether the underlying causal or statistical assumptions for the fuzzy RD design are warranted. Finally, we demonstrate our approach by studying the effect of extended time accommodations using data from the National Assessment of Educational Progress.

Regression Discontinuity

Geographic Regression Discontinuity Design with Spatially Correlated Noise Jiayin Li* Jiayin Li, David Ruppert, Y. Samuel Wang,

This paper introduces a Bayesian nonparametric method tailored for Geographic Regression Discontinuity Design (GeoRDD) to address causal inference challenges in spatial settings, particularly under spatially correlated noise. Traditional Regression Discontinuity Design (RDD), which assesses causal effects by comparing outcomes near a cutoff, is expanded to GeoRDD, using geographic boundaries as forcing variables. Our method, utilizing Gaussian process regression, addresses the limitations of models that assume independent, identically distributed (i.i.d.) noise, which are often overly simplistic for spatial contexts. Through a case study examining the influence of school district boundaries on housing prices in New York City, we demonstrate the practical application and relevance of our proposed method and illustrate its accuracy and reliability in estimating causal effects in spatial contexts. The findings emphasize the enhancement of causal effect estimation achieved by integrating spatially correlated noise and highlight the importance of considering spatial interdependencies in GeoRDD analyses.

Sensitivity Analysis

Unifying L_2 sensitivity analyses for regression and weighting estimators Yaxuan Huang*
Yaxuan Huang, Melody Huang, Samuel Pimentel,

Sensitivity analyses addressing possible unobserved confounding are a vital piece of evidence when causal inferences are drawn from observational studies. Recent methodological work has introduced many new approaches for sensitivity analysis; however, each method is typically tied to a particular estimation strategy, making it difficult to compare sensitivity approaches across studies. Leveraging recent work on connections between regression estimators and weighting estimators (Chattopadhyay & Zubizarreta 2023), we establish a new unified framework for sensitivity analysis for both regression and weighting, under which standard approaches for both types of estimators appear as alternative parameterizations of an L_2 constraint on the errors from unobserved confounding. We show how our framework can lead to intuitive characterizations of L_2 sensitivity analysis for doubly robust and weighted regression estimators of causal effects, and allows bounds to be made sharp without estimating additional nuisance parameters. We discuss conceptual parallels of our framework relative to recently-introduced doubly valid, doubly sharp L_infty sensitivity analyses. By moving away from worst-case characterizations of unobserved confounding error, the proposed L_2 approach results in improvements in both stability and interpretability.
Supported by the National Science Foundation under Grant No. 2142146.

Sensitivity Analysis

Sensitivity Analysis with Likelihood Ratio Test and Pearson's chi-square Test from IxJ Tables Elaine Chiu* Elaine Chiu, Hyunseung Kang,

Examining associations between categorical variables via contingency tables is common in clinical and social science research. Typically, the strength of these associations is measured using a Likelihood ratio test (LRT) or Pearson's chi-square test. However, in observational studies, these associations do not imply causation due to unmeasured confounding and a sensitivity analysis seeks to understand how the associations can be nullified by an unmeasured confounder. This paper proposes a non-asymptotic, exact sensitivity analysis for tests of associations in IxJ contingency tables. In particular, we extend the Rosenbaum sensitivity model to allow for (a) non-binary, potentially un-ordered, exposures, (b) a larger class of test statistics, and (c) two-sided alternatives typically implied in an LRT or the Pearson's chi-square test. We apply our method to assess the association between three types of pre-kindergarten (pre-k) care and students' overall math achievement, measured on a discrete scale, from the Early Childhood Longitudinal Study-Kindergarten cohort. After controlling for socioeconomic and demographic factors, we find that the association between pre-k programs and math performance is strong, especially for black and Hispanic female students (two-sided p-values: 0.0034 and 0.0086 when $\gamma=0$), and the observed association is insensitive up to a Rosenbaum's Γ of 2.

Sensitivity Analysis

A Split-Sampling Framework for Powerful Design of Observational Studies under Unmeasured Confounding William Bekerman* William Bekerman, Abhinandan Dalal, Dylan Small,

Observational studies are valuable tools for inferring causal effects in the absence of controlled experiments. However, these studies may be biased due to the presence of some relevant, unmeasured set of covariates. The design of an observational study has a prominent effect on its sensitivity to hidden biases and the best design may not be apparent without examining the data. One approach to facilitate a data-inspired design is to split the sample into a planning sample for choosing the design and an analysis sample for making inferences. This procedure has been shown to enhance power when it is assumed that at most one among multiple outcomes is affected by the treatment and a single outcome is chosen in the planning sample. We devise a powerful and flexible method for selecting outcomes in the planning sample when more than one outcome may be affected by the treatment. We investigate the theoretical properties of our method and conduct extensive simulations that demonstrate pronounced benefits, especially at higher levels of allowance for unmeasured confounding. Finally, we demonstrate our method in an observational study of the multi-dimensional impacts of a devastating flood in Bangladesh.

Sensitivity Analysis

Sensitivity analysis with multiple treatments and multiple outcomes with applications to air pollution mixtures Joseph Antonelli* Joseph Antonelli, Suyeon Kang, Alexander Franks,

Understanding the health impacts of air pollution is vital in public health research. Numerous studies have estimated negative health effects of a variety of pollutants, but accurately gauging these impacts remains challenging due to the potential for unmeasured confounding bias that is ubiquitous in observational studies. In this study, we develop a framework for sensitivity analysis in settings with both multiple treatments and multiple outcomes simultaneously. This setting is of particular interest because one can identify the strength of association between the unmeasured confounders and both the treatment and outcome, under a factor confounding assumption. This provides informative bounds on the causal effect leading to partial identification regions for the effects of multivariate treatments that account for the maximum possible bias from unmeasured confounding. We also show that when negative controls are available, we are able to refine the partial identification regions substantially, and in certain cases, even identify the causal effect in the presence of unmeasured confounding. We derive partial identification regions for general estimands in this setting, and develop a novel computational approach to finding these regions.

Sensitivity Analysis

Creating Control Groups for Assessing Market-Level Treatments: Assessing a Promotional Campaign Aimed at Increasing Sales Stella McMullen* Stella McMullen,

Estimating the impact of product rollouts, price changes, or marketing campaigns is crucial for companies aiming to make informed decisions. However, conducting individual-level experiments can be challenging or unfeasible. An alternative approach is implementing changes in one market and comparing outcomes to other markets. This paper focuses on providing best practices for creating suitable control groups specifically for market-level treatments.

In the context of evaluating a promotional campaign, we present a comprehensive overview of methodologies, an overview of results, sensitivity analysis considering the impact of model choice, covariate selection, and estimation period on the selected control, and general guidelines for identifying a control group. The methodologies discussed include matching techniques tailored for time-series cross-sectional data and the Synthetic Control Method.

This research is valuable for assessing the effectiveness of business changes and is relevant to researchers and practitioners interested in market-level interventions. It contributes to the field’s understanding of selecting appropriate control groups for market-level treatments and sensitivity to the modeling process. The findings provide actionable insights that inform evidence-based decision-making and strategic planning.

Synthetic Control Method

Single Proxy Synthetic Control Chan Park* Chan Park, Eric Tchetgen Tchetgen,

Synthetic control methods are widely used to estimate the treatment effect on a single treated unit in time-series settings. A common approach for estimating synthetic controls is to regress the treated unit's pre-treatment outcome on those of untreated units via ordinary least squares. However, this approach can perform poorly if the pre-treatment fit is not near perfect, whether the weights are normalized or not. In this paper, we introduce a single proxy synthetic control approach, which views the outcomes of untreated units as proxies of the treatment-free potential outcome of the treated unit, a perspective we leverage to construct a valid synthetic control without the need for an interactive fixed effect model or a perfect pre-treatment fit. Under this framework, we establish alternative identification and estimation methodologies for synthetic controls and for the treatment effect on the treated unit. Notably, unlike a proximal synthetic control approach which requires two types of proxies for identification, ours relies on a single type of proxy, thus facilitating its practical relevance. Additionally, we adapt a conformal inference approach to perform inference about the treatment effect, obviating the need for a large number of post-treatment data. Lastly, our framework can accommodate time-varying covariates and nonlinear models. We demonstrate the proposed approach in a simulation study and a real-world application.

Synthetic Control Method

Inference for Synthetic Controls via Refined Placebo Tests Timothy Sudijono* Timothy Sudijono, Lihua Lei,

The synthetic control method is often applied to problems with one treated unit and a small number of control units. Inference procedures that are justified asymptotically are often unsatisfactory due to (1) small sample sizes that render large-sample approximation fragile and (2) simplification of the estimation procedure that is actually implemented in practice. An alternative is design-based inference, which is closely related to the placebo test, a widely used diagnostic tool in practice. It provides valid Type-I error control in finite samples without artificial simplifications of the method when the treatment is assigned uniformly among units. Despite this robustness, it suffers from low resolution since the null distribution is constructed from only N reference estimates, where N is the sample size. Inspired by a connection to the conformal inference literature, we propose a novel leave-two-out procedure that bypasses this issue, providing $O(N^2)$ reference estimates while still maintaining finite-sample Type-I error control under uniform assignments. Unlike the placebo test whose Type-I error always equals the theoretical upper bound, our procedure often achieves a lower Type-I error than theory suggests and a higher power when the effect size is reasonably large. To account for deviation from uniform assignments, we generalize our procedure to allow for non-uniform assignments and show how to conduct sensitivity analysis based on quadratic programming.

Synthetic Control Method

A new perspective on synthetic controls Yujin Jeong* Yujin Jeong, Dominik Rothenhäusler,

In statistics and machine learning, we often want to quantify uncertainty and frame optimality with respect to sampling uncertainty. However, if we combine evidence from different data sets, sampling uncertainty might be lower order than the distribution shift between the data sets. This raises the question of how to optimally estimate in a data fusion setting. To address this issue, we model distributional shifts as a superposition of numerous random changes. We then develop tools for measuring the similarity between randomly perturbed distributions, estimating parameters of perturbations, and predicting outcomes for new distributions. Interestingly, these tools share a close connection to synthetic controls. Our framework provides a new language for distributional shifts and offers a fresh perspective on synthetic controls. Moreover, we evaluate its performance on real-world data sets and demonstrate that our new language and tools significantly improve estimation accuracy.

Synthetic Control Method

The Perils of Nonstationary Data in Synthetic Control Applications Hongyu Mou* Hongyu Mou, Yiqing Xu, Ziyi Liu, Yifan Sun,

The synthetic control method (SCM) is widely used to estimate treatment effects in comparative case studies. However, inference with SCM remains challenging. Existing inferential approaches include randomization inference, which requires random assignment of the treatment, or conformal inference, which demands stationary or cointegrated error term time-series. While justifying random assignment is often difficult, our replication of thirteen SCM applications in economics and political science shows that the stationarity requirement is frequently unmet. Rosenbaum-type sensitivity analyses or analyses based on stationarized data indicate that many existing SCM findings are either spurious or underpowered. To obtain valid inference, we recommend a programmatic diagnostic procedure for future SCM applications.

Target Trial emulation / Application of causal inference

Emulation of a target trial to estimate the per-protocol effect of switching from triple to dual antiretroviral therapy on clinical outcomes in people with HIV Sophia Rein* Sophia Rein,

Background
Intention-to-treat analyses of randomized trials found similar rates of virologic failure in people with HIV switching from triple to dual antiretroviral therapy (ART) compared with those continuing on triple ART. We emulated a target trial to estimate the per-protocol effect of switching from triple to dual ART on virologic failure.
Methods
To emulate the target trial, we identified eligible individuals in 10 observational cohorts between January 2015 and April 2023. Using a sequential emulation with a potential time zero for each individual during each month, we estimated the 5-year risk of virologic failure via pooled logistic regression with censoring if individuals were nonadherent to their initial strategies. We used non-stabilized inverse probability weights to adjust for baseline and time-varying confounding, and nonparametric bootstrapping with 500 samples to calculate 95% confidence intervals (CIs).
Results
During follow-up, adherence to the initial regimen was about 80% in 2,921 included dual therapy initiators and 76% in 79,565 non-initiators. The estimated 5-year risks (95%CI) of virologic failure were 4.38% (3.06-5.68) in dual therapy initiators and 4.19% (3.95-4.47) in non-initiators (risk ratio: 1.04 (0.74-1.35)).
Conclusion
We estimated that switching from triple to dual ART had little impact on virologic failure. Analyses on differences in the risk of clinical events and death, where evidence from randomized trials is lacking, are ongoing.

Weighting

Bridging Binarization: Causal Inference with Dichotomized Continuous Treatments Kaitlyn Lee* Kaitlyn Lee, Alejandro Schuler, Alan Hubbard,

The average treatment effect (ATE) is a common parameter estimated in causal inference literature, but it is only defined for binary treatments. Thus, despite concerns raised by some researchers, many studies seeking to estimate the causal effect of a continuous treatment create a new binary treatment variable by dichotomizing the continuous values into two categories. In this paper, we affirm binarization as a statistically valid method for answering causal questions about continuous treatments by showing the equivalence between the binarized ATE and the difference in the average outcomes of two specific modified treatment policies. These policies impose cut-offs corresponding to the binarized treatment variable and assume preservation of relative self-selection. Through this equivalence, we clarify the assumptions underlying binarization and discuss how to properly interpret the resulting estimator. Additionally, we introduce a new target parameter that can be computed after binarization that considers the status-quo world. We argue that this parameter addresses more relevant causal questions than the traditional binarized ATE parameter. Finally, we present a simulation study to illustrate the implications of these assumptions when analyzing data and to demonstrate how to correctly implement estimators of the parameters discussed.

Weighting

Poisson regression under heterogeneous treatment effects Georgy Kalashnov* Georgy Kalashnov, Lihua Lei,

We study the properties of Poisson regression under heterogeneous treatment effects as a better choice compared to log-linear regression in policy evaluation setting. Log-linear regression aggregates multiplicative individual effects from both large and small subjects with equal weights. However, a policy evaluation study is interested in a ratio of the means rather than mean of the ratios, which is to give weights proportional to the subject sizes. We show that Poisson regression estimates exactly the ratio of means in different specifications. We show that a Poisson regression on treatment and controls estimates a convex average of the individual effects in the spirit of Angrist (1998) in the case when the effects are small and average treatment on control, when the effects are large and positive, and average treatment on the treated, when the effects are small. We also show a double robust way to estimate an average treatment effect in Poisson regression.

Bayesian Causal Inference

Extending general BART with Pitman-Yor mixtures: novel nonparametric prior to correct for strong unobserved confounding Andrej Srakar* Andrej Srakar, Marilena Vecco,

Bayesian additive regression trees (BART) perspective has been developed by Chipman et al. (2010) and popularized in its usage in causal inference problems. It commonly uses a specific regularization prior, sometimes combined with Gaussian, Dirichlet, Dirichlet Process Mixture and semiparametric perspectives in a general BART perspective (Tan and Roy, 2019). Despite its success there has been a growing number of papers that point out its limitations. We develop a novel nonparametric regularization prior for BART based on Pitman-Yor Mixture (PYM) partition-based process standard error structure, which has to date to our knowledge rarely been used in causal inference. Our novel perspective is studied for several causal perspectives: regression discontinuity design; causal maximally partially directed acyclic graph; direct causal clause; and causal mediation. Study of asymptotic properties in a Bayesian framework extends recent proposal of Jeong and Rockova (2022) of sparse piecewise heterogeneous anisotropic Hölder functions to account for anisotropic regions in general BART. Results on simulated and real data confirm improved properties compared to earlier BART priors in particular in the presence of strong confounding. We address computational issues by using importance sampling with the integrated nested Laplace approximation (Outzen Berild et al., 2021). We discuss extensions to endogeneity corrections and Single World Intervention Graph perspective.

Bayesian Causal Inference

Optimizing Clinical Trial Design with Causal Learning: Insights and Predictions for Improved Success Rates Shaurya Gaur* Shaurya Gaur, Lucia Pagani, Lorenzo Rigolli,

Clinical trials, crucial for evaluating new treatments, are becoming increasingly expensive and time-consuming with a substantial number of trials failing due to intrinsic design challenges. In this paper, we propose an optimization tool for adjusting clinical trial design to increase their probability of success. Our tool utilizes a weighted logistic regression predictive model trained on a dataset of 41,269 trials from clinicaltrials.gov (CTGov), with labels obtained from TrialTrove.
The weights enable us to decorrelate features and obtain a causal interpretation of their impact on trial outcomes. Given the parametric nature of logistic regression, we can easily maximize the probability of trial success within a specific domain, allowing us to identify and implement optimized designs for proposed trials.
The AUROC score obtained by the logistic regression predictive model is 0.623. Furthermore, we collected evidence on the improved quality of optimized designs by comparing scores obtained from non-linear models. The best-performing model achieved a score of 0.695; however, it is less interpretable compared to our model.
The practical implications of this work are significant for pharmaceutical companies aiming to enhance their trial success rates. By employing our tool, they can optimize trial designs, potentially saving resources and increasing the likelihood of successful outcomes.
Keywords: Clinical trials, Weighted Logistic Regression, Causal learning, Optimization

Bayesian Causal Inference

Regression discontinuity under clustering Kevin Tao* Kevin Tao,

The Regression Discontinuity Design (RDD) is a popular quasi-experiment for estimating the local average treatment effect when randomization can not be performed. Despite the abundance of literature on RDD, few have considered the scenario where sub-populations with heterogeneous treatment effects are present. We generalize the existing RDD framework by considering the presence of sub-populations with different conditional mean functions, effectively extending from the one parameter estimation problem of RDD, to simultaneous estimation of multiple treatment effect on the same domain. We propose the method of Hierarchical Gaussian Process Regression (HGPR). Our method not only enjoys the flexibility of regular GPR, but also allows for correlation between sub-populations, and most importantly enable us to pool information across sub-populations. We derive the posterior distribution, perform a Bayes risk analysis, and propose a Metropolis-Hasting within Gibbs algorithm for fitting our HGPR. Finally, we compare our method to local linear regression, the state-of-art method, via simulation studies.

Causal Discovery

Causality Inspired Models for Trading-off Invariance and Prediction Error in Financial Time Series Forecasting Xi Lin* Yutong Lu, Daniel Cunha Oliveira, Xi Lin,

Time series forecasting in finance is a pivotal task with significant implications for investment strategies, risk management, and economic planning. However, it is fraught with challenges due to the inherent complexity, noise, and volatility of financial markets. Conventional forecasting models often fail to generalize when faced with regime switching and distributional shifts. In this research, we leverage the use of causal discovery and invariant prediction techniques to resolve the aforementioned obstacles in asset returns forecasting.

We introduce a novel framework which integrates causal discovery, to identify causal predictors, with forecasting models. This approach balances the trade-off between invariance to distributional changes and minimization of prediction errors. To the best of our knowledge, we are the first to conduct comparative analysis among state-of-the-art causal discovery algorithms, for example LiNGAM, DYNOTERS, Invariant Causal Prediction, etc., benchmarked against non-causal feature selection techniques, in the application of forecasting asset returns. Our empirical evaluations demonstrate the efficacy of our approach in yielding stable and accurate predictions, outperforming baseline models, particularly in tumultuous market conditions.

Causal Discovery

Causal discovery with expert knowledge Aparajithan Venkateswaran* Aparajithan Venkateswaran, Emilija Perkovic,

We consider Markov equivalence classes of maximal ancestral graphs (MAGs) and their restrictions. MAGs are directed mixed graphs used to model conditional independence constraints between a set of observed variables which correspond to nodes in the graph. All MAGs that imply the same set of conditional independence relationships form a Markov equivalence class, which can be uniquely represented by a partially oriented graph (essential graph). Past work has seen the development of algorithms for learning an essential graph from observational data. Recently, there has been interest in learning restrictions of the Markov equivalence class under specific kinds of causal background knowledge. In this work, we focus on restrictions of the Markov equivalence class formed by fixing certain edgemarks.
First, we generalize a property previously formalized by Zhao et al. [2005] and prove a conjecture by Ali et al. [2009] for MAGs in a Markov equivalence class. Second, we incorporate edgemark background knowledge into an essential graph analogous to Meek [1995]. We prove soundness for two new orientation rules in addition to the previously established rules of Spirtes et al. [1999] and Zhang [2008]. We further refine some of the orientation rules of Zhang [2008] in our setting. Finally, we provide a sound algorithm to restrict the equivalence class of MAGs using background knowledge. We show that our algorithm is complete for a special case and conjecture that it is true for all graphs.

Causal Discovery

Post-selection inference for causal effects after causal discovery Ting-Hsuan Chang* Ting-Hsuan Chang, Zijian Guo, Daniel Malinsky,

Algorithms for constraint-based causal discovery select graphical causal models from among a space of possible candidates (e.g., all directed acyclic graphs) by executing a sequence of conditional independence tests. These may be used to inform the estimation of causal effects (e.g., average treatment effects) when there is uncertainty about which covariates ought to be adjusted for, or which variables act as confounders versus mediators. However, naively using the data twice, for model selection and estimation, would lead to invalid confidence intervals. Moreover, if the selected graph is incorrect, the inferential claims may apply to a chosen functional that is distinct from the actual causal effect. We propose an approach to post-selection inference that is based on a resampling procedure, that essentially performs causal discovery multiple times with randomly varying intermediate test statistics. Then, an estimate of the target causal effect and corresponding confidence sets are constructed from a union of individual graph-based estimates and intervals. We show that this construction has asymptotically correct coverage. Though most of our exposition focuses on the PC algorithm for learning directed acyclic graphs and the multivariate Gaussian case for simplicity, the approach is general and modular, so it can be used with other conditional independence based discovery algorithms and (semi-)parametric families.

Causal Discovery

Decorrelation of Dependent Discrete Data for Causal Discovery Alex Chen* Alex Chen, Qing Zhou,

The assumption of independence between observations is prevalent across various methodologies for causal graph estimation. However, this assumption does not always hold in practice, posing challenges to accurate structure learning. To address this, we propose a Discrete Decorrelation algorithm (DDA) for causal graph learning on dependent data, where the local conditional distribution is defined by a latent utility model with dependent error variables. We propose a pairwise likelihood method to estimate the covariance matrix for the dependence among the units. Leveraging the estimated covariance matrix, we develop an EM-like iterative algorithm to generate and de-correlate samples of the latent utility variables, which serve as de-correlated datasets. Then, any standard causal discovery method can be applied on the de-correlated data to learn the underlying causal graph. Our method enables the application of traditional causal discovery methods, which were developed for independent data, to dependent data. We demonstrate that the proposed method significantly improves the accuracy in causal graph learning, through evaluations on both synthetic and real-world datasets.

Causal Inference and Bias/Discrimination

The Functional Average Treatment Effect Shane Sparkes* Shane Sparkes, Lu Zhang, Erika Garcia,

This paper establishes the functional average as an important estimand for causal inference. The significance of the estimand lies in its robustness against traditional issues of confounding. We prove that this robustness holds even when the probability distribution of the outcome, conditional on treatment or some other vector of adjusting variables, differs almost arbitrarily from its counterfactual analogue. This paper also examines possible estimators of the functional average, including the sample mid-range, and proposes a new type of bootstrap for robust statistical inference: the Hoeffding bootstrap. After this, the paper explores a new class of variables, the U class of variables, that simplifies the estimation of functional averages. This class of variables is also used to establish mean exchangeability in some cases and to provide the results of elementary statistical procedures, such as linear regression and the analysis of variance, with causal interpretations. Simulation evidence is provided. The methods of this paper are also applied to a National Health and Nutrition Survey data set to investigate the causal effect of exercise on the blood pressure of adult smokers.

Causal Inference and Bias/Discrimination

A Decision-Theoretic Framework for Sample Selection in Randomized Experiments Yuchen Hu* Yuchen Hu, Stefan Wager, Emma Brunskill, Henry Zhu,

The design of randomized experiments often fails to account for heterogeneity of treatment effects across different subpopulations, and discussions on how to reflect different fairness-oriented desiderata in study design are largely absent from the literature. For example, until recently, most medical research in the United States was conducted on white men, while excluding women and racial minorities (e.g., Dresser, 1992, “Wanted single, white male for medical research”); and FDA-approved trials still under-sample black participants relative to their share of the population (Alsan, EAAMO, 2022).

While numerous studies have explored treatment assignment strategies for a given sample, there has been limited discussion on the initial selection of a sample from a heterogeneous population. To address these issues, we study how various decision-theoretic frameworks, including minimax regret, utility maximization and cooperative bargaining, can be used to guide sample selection in randomized experiments. We consider a model where different subpopulations may differentially benefit from the knowledge gained in the study, and study participation may involve burdens or rewards which may also manifest themselves differentially among groups. We illustrate how different beliefs and objectives can lead to notably different sample allocations.

Causal Inference and Bias/Discrimination

Regression-Based Proximal Causal Inference Jiewen Liu* Jiewen Liu, Chan Park, Kendrick Li, Eric Tchetgen Tchetgen,

A recently proposed framework to account for known but unmeasured sources of confounding is so-called proximal causal inference (PCI). The approach leverages negative controls, more broadly termed treatment and outcome confounding proxies, a priori known to have null associations with the primary treatment and outcome, respectively, conditional on measured and hidden confounders which they proxy. While formal statistical inference has been developed for PCI, its implementation is hindered by complex, ill-posed integral equations. Our paper introduces a regression-based PCI approach, obviating solving integral equations. Our method employs two-stage regression through generalized linear models (GLMs). In the first stage, one fits a GLM for an outcome proxy in terms of the treatment proxy. In the second stage, one fits a GLM for the primary outcome using the predicted value of the first stage regression model as a regressor that accounts for residual confounding. The proposed approach has merit in that (i) it is applicable to continuous, count, and binary outcomes, making it relevant to a wide range of real-world applications, and (ii) it is easy to implement by using off-the-shelf software. We establish statistical inference theory for regression-based PCI and illustrate their performance in both synthetic and real-world empirical applications.

Causal Inference and Bias/Discrimination

Pseudo-Robust Solutions to Lord’s Paradox Robert Larzelere* Robert Larzelere, Hua Lin,

Analyses of residualized vs. simple change scores often produce contradictory treatment estimates in non-randomized longitudinal studies, a problem known as Lord’s [, 1967, Lord] Paradox. Under some assumptions, the two estimates bracket the unbiased causal effect [Angrist, 2009, Pischke], and the econometric practice of testing robustness across contrasting analyses has been recommended [Duncan, 2014]. Unfortunately, robust consistency across analyses of residualized and simple difference scores often occurs artifactually, which we call pseudo-robustness. We present four pseudo-robust solutions to Lord’s Paradox: (1) Making the pretest a covariate in difference-score analyses makes their treatment effects identical to ANCOVA’s treatment effects. (2) Pretest matching makes the two treatment effects equal to each other and to ANCOVA’s treatment effect before matching, whether biased or not. (3) Centering pretest and posttest scores on pretest group means also produces equivalent treatment effects, but ones equivalent to the treatment effect of difference-score analysis before centering, whether biased or not. (4) Two treatments for depression still appear robustly harmful in secondary analyses of the most at-risk of three subgroups even after propensity-score matching. We need to distinguish appropriate robustness and causal-estimate bracketing from pseudo-robustness in these 2-occasion analyses and in more complex longitudinal analyses.

Causal Inference and Bias/Discrimination

Sampling based on Milestones (SMile): A potential alternative design Ian Shrier* Ian Shrier, Tibor Schuster, Yi Li, Zachary Vernec, Russell Steele,

Some authors suggest simulating data to evaluate potential biases prior to starting a study. We wanted to evaluate a current hypothesis that concussion causes symptoms through decreases in binocular vision near point convergence (NPC). Because symptom resolution ranges from days-months, sampling at fixed timepoints is inefficient. Rather, if the NPC vs. symptom relationship is time-independent, we expect no bias with sampling based on milestones (SMile design): time of concussion, 50% improved symptoms, and healed. We created Oracle data (comprehensive Monte Carlo simulation study) using the same healing rate and linear slope for NPC vs. symptom for each participant. Bias was the difference of the average slope and the Oracle data slope (= 1). There was never bias when there was no measurement error. With multiplicative measurement error up to 50%, the overall bias was <5%. Grouped by initial symptom score, the slope was underestimated by ~30% for the mild (low symptom score) group, and overestimated by ~8% for the severe group (high symptom score). Bias increased in the mild group when we excluded participants who healed quickly. Bias decreased when measurement was delayed by days after the milestone. All these biases completely depend on observed error measurement. The SMile design may be useful when, within the same individual, two linearly related characteristics (ie. NCP and symptom scores) can be precisely measured over time. Further work will explore other assumptions.

Causal Inference and Bias/Discrimination

Proximal Causal Inference with Some Invalid Proxies Prabrisha Rakshit* Prabrisha Rakshit, Eric Tchetgen Tchetgen,

In observational studies, researchers concerned with confounding due to hidden factors have recently adopted proximal causal learning to identify and estimate causal effects. The approach acknowledges that even in well-designed observational studies, covariate measurements may at best be proxies of underlying confounding mechanisms. But traditional proximal causal learning relies on prior knowledge of the validity and relevance of proxies : a valid treatment inducing proxy should not directly affect the outcome and a relevant outcome inducing proxy must be related to treatment only to the extent that it is related to an unmeasured confounder. However obtaining complete knowledge about the validity of proxies may often be impractical. This paper introduces necessary-sufficient conditions to identify a causal effect when such a priori knowledge is lacking. We propose a 2-stage estimator based on the LASSO for estimating a causal effect, with theoretical guaranties regarding its performance. We further suggest a 2-stage adaptive LASSO-based estimator for the causal effect, incorporating adaptive weights to penalize different coefficients in the L1 penalty. This is to address scenarios in which LASSO exhibits inconsistency in variable selection. We show that this estimator possesses oracle properties. We further correct the bias introduced by L1 penalization and establish the limiting distribution of the debiased estimator for the causal effect.

Causal Inference and Bias/Discrimination

Efficient and doubly robust estimation of COVID-19 vaccine effectiveness under the test-negative design Cong Jiang* Cong Jiang,

While the test-negative design (TND), which is routinely used for monitoring seasonal flu vaccine effectiveness (VE), has recently become integral to COVID-19 vaccine surveillance, it is susceptible to selection bias due to outcome-dependent sampling. Some studies have addressed the identifiability and estimation of causal parameters under the TND, but efficiency bounds for nonparametric estimators of the target parameter under the unconfoundedness assumption have not yet been investigated. We propose a one-step doubly robust and locally efficient estimator called TNDDR (TND doubly robust), which utilizes sample splitting and can incorporate machine learning techniques to estimate the nuisance functions. We derive the efficient influence function (EIF) for the marginal expectation of the outcome under a vaccination intervention, explore the von Mises expansion, and establish the conditions for $sqrt{n}-$consistency, asymptotic normality and double robustness of TNDDR. The proposed TNDDR is supported by both theoretical and empirical justifications, and we apply it to estimate COVID-19 VE in an administrative dataset of community-dwelling older people (aged $geq 60$y) in the province of Québec, Canada.

Causal Inference and Bias/Discrimination

An Individual Causal Framework for Evaluating Electoral Systems Cory McCartan* Cory McCartan, Christopher Kenny,

Social scientists have developed numerous criteria for evaluating different electoral systems. But these largely depend on aggregate features of the electoral system or its results, not on individual outcomes. Because of this, they are not easily adapted to quantify the impacts of electoral systems on individuals or on groups of individuals defined by race, class, or geography. Moreover, they are ill-suited to capture the effects of changes between electoral systems. To overcome these limitations, we propose a unified causal framework for measuring the effects of electoral reforms on individuals: who benefits, who is harmed, where they live, and what groups they belong to. We define causal measures that zero in on voters whose representational outcomes change as a result of the electoral reform, for better or for worse, and can aggregate these individual gains and losses to quantify differential effects on various groups of voters. The framework and proposed measures improve on existing approaches by both focusing on the choices of individual voters and directly incorporating counterfactual electoral systems, which are always relevant in reform settings. We discuss identification and estimation strategies, and demonstrate the utility of our framework through analyses of voting rights litigation in Alabama, the adoption of ranked-choice voting in Alaska, and redistricting criteria changes in Washington.

Causal Inference and SUTVA/Consistencies Violations

Casual Inference in the Presence of Limited Overlap Michael Elliott* Michael Elliott, Tingting Zhou, Rod Little,

Propensity score (PS) based methods are often used to control for observed confounders in observational studies of causal effects. For PS methods to work reliably, there should be sufficient overlap in the propensity score distributions between treatment groups. Limited overlap can result in fewer subjects being matched or in extreme weights causing numerical instability and bias in causal estimation. The problem of limited overlap suggests (a) defining alternative estimands that restrict inferences to subpopulations where all treatments have the potential to be assigned, and/or (b) excluding or down-weighting sample cases where the propensity to receive one of the compared treatments is close to zero. We compared several PS methods for estimation of alternative causal estimands when limited overlap occurs. Simulations suggest that, when there are extreme weights, penalized spline of propensity prediction that we recently developed tends to outperform the weighted estimators for ATE and performs similarly to the weighted estimators for alternative estimands. We illustrate with an example that assess whether right heart catheterization (RHC) is a beneficial treatment in treating critically ill patients.

Causal Inference and SUTVA/Consistencies Violations

Estimation and Inference under Recommender Interference Ruohan Zhan* Ruohan Zhan, Shichao Han, Yuchen Hu, Zhenling Jiang,

Recommendation algorithms are crucial in digital platforms for tailoring content to viewer preferences. These platforms rely on content creators to maintain a dynamic viewer community. This study evaluates interventions targeting creators. Using canonical creator-side randomization A/B experiments, we find that the standard difference-in-mean estimator is biased in estimating treatment effects. This bias stems from the interference among creators competing for visibility through recommendation algorithms. To address this, we propose an innovative method to eliminate interference bias in measuring treatment effects. We introduce a semi-parametric model for recommender choice and develop influence functions for treatment effects that satisfy Neyman orthogonality. This approach enables us to create a consistent and asymptotically normal estimator for treatment effects, supporting inference and hypothesis testing. We demonstrate the efficacy of our method through simulations and practical applications on a leading short video platform.

Causal Inference and SUTVA/Consistencies Violations

Outcome Modeling in Design-Based Inference for Spatial Experiments Arisa Sadeghpour* Arisa Sadeghpour, Erin Hartman,

In many spatial settings, randomized treatments have effects that bleed out, violating the standard non-interference assumption, and researchers often want to estimate how these spillover effects decay in space. Wang et al. (2023) present a design-based estimand for the average marginalized effect at a specific distance, which we call the “circle average marginalized effect” (CAME). The CAME is the average effect of switching intervention nodes from treatment to control on points along a circle a specific distance away from those nodes, marginalizing over possible realizations to other nodes. Since it is impossible to observe outcomes at every point in space, we argue that estimating CAME necessitates using outcome models. Through simulations, we explore how the choice of outcome model impacts the bias and variance of the Horvitz-Thompson and Hajek estimators for the CAME in different scenarios. We find that when using modeled outcomes as in practice, even the Horvitz-Thompson estimator accrues some bias, under design-based inference, as a result of the modeling. We illustrate the importance of outcome modeling in estimating the CAME using randomized field experiments on policing and crime.

Causal Inference and SUTVA/Consistencies Violations

Symbiosis bias in A/B tests of Recommendation Algorithms David M. Holtz* Jean Pouget-Abadie, Jennifer Brennan, David M. Holtz,

One assumption underlying the unbiasedness of global treatment effect estimates from randomized experiments is the stable unit treatment value assumption (SUTVA). Experiments that compare the efficacy of different recommendation algorithms often violate SUTVA, because each algorithm is trained on a pool of shared data produced by the different recommendation algorithms being evaluated in the experiment. This shared training data across recommendation algorithms can lead to serious bias in the routine evaluations of such algorithms. We illustrate the presence and magnitude of this bias, which we call “symbiosis bias,” in a real data study on a large tech platform. We further explore, through simulation, cluster randomized and data-diverted solutions to mitigating this bias, and use a stylized analytical model to characterize the relative efficacy of these two solutions at reducing symbiosis bias under different conditions.

Causal Inference and SUTVA/Consistencies Violations

Low-degree Outcomes and Clustered Designs: A Combined Approach for Causal Inference under Interference Matthew Eichhorn* Samir Khan, Johan Ugander, Matthew Eichhorn, Christina Lee Yu,

One line of work in causal inference under interference develops estimators that are low variance under parametric outcome models; another develops experimental designs that reduce variance under assumptions on the interference graph. Recent work on low-degree outcome models and pseudoinverse (PI) estimators exemplifies the former, while work on graph cluster randomization exemplifies the latter. In this work, we explore the intersection of these two by studying the interplay between low-degree models and clustered designs.

We extend the analysis of PI estimators for low-degree models beyond Bernoulli designs, characterizing unbiasedness and giving variance bounds. For clustered designs, we show that the variance of the PI estimator scales like the minimum of the variance bound guaranteed by either cluster randomization or low-degree modeling on their own. Thus, the PI estimator has consistently less variance than the Horvitz-Thompson (HT) estimator when the clustering is fixed, a fact we verify empirically.

In contrast, for randomized clustered designs, which were based on the variance structure of the HT estimator, we find that the PI estimator may have higher variance than the HT estimator. Thus the PI estimator is appropriate when using clustered designs, but not necessarily when using randomized clustered designs; as such our results provide initial answers to open questions about the alignment between design and estimator in causal inference under interference.

Causal Inference and SUTVA/Consistencies Violations

Non-Existent Outcomes in Research on Inequality: A Causal Approach Ian Lundberg* Ian Lundberg, Soonhong Cho,

When studying inequality, a focal outcome may not exist for some individuals. Those who are not employed have no hourly wage, for example. Scholars of wage inequality routinely drop the non-employed. But the same causal process that shapes wage inequality among the employed also shapes which people are employed at all. Researchers who drop those with non-existent outcomes inadvertently induce selection problems and obscure inequality. We show how to use principal stratification methods to study two quantities: (1) the average effect on whether an outcome exists, and (2) the average effect on that outcome among the latent set of people who would have an outcome under either treatment condition. Our technical contribution is to carry out principal stratification within a parametric regression analysis that adjusts for measured confounders. Our applied contribution is to reveal how standard practices in sociology and economics obscure inequality. We illustrate by showing how past work has understated the causal effect of motherhood on the hourly wages of women who would be employed with or without children.

Causal Inference and SUTVA/Consistencies Violations

Staggered Difference-in-Differences with Partial Interference Anni Hong* Anni Hong, Eli Ben-Michael,

Difference-in-Differences (DiD) designs are a workhorse approach for modern policy evaluation. However, most existing methods rule out treatment effect spillovers and interference between units. Spillovers and interference are likely in many settings, especially when units are situated geographically or within social networks. In such cases, the impact of a policy can be affected by which and how many units are treated. We develop a framework for DiD under staggered adoption of treatment and interference that addresses this challenge by exploiting partial interference within disjoint clusters, such as neighborhoods or classrooms, without requiring knowledge of the underlying structure of interference within clusters. We show how to identify, estimate, and construct asymptotically valid confidence intervals for a family of treatment effect estimands within this setting, allowing for both the number of clusters and the cluster sizes to grow. We find that the asymptotic rates of convergence depend on both the amount of correlation within clusters and the number of independent clusters. Furthermore, we extend this framework and show how to leverage staggered adoption of treatment when there is limited, but still unknown, interference and temporal spillovers to extrapolate effects and explore scenarios where everyone is treated at a specific time. We demonstrate the utility of the proposed methods with extensive simulation studies and a real-world example.

Causal Inference Education

Online Learning in the Face of Unemployment Kiet Le* Kiet Le, Octavio Aguilar,

Shortly after the onset of the COVID-19 pandemic the unemployment rate reached 14.7%–the highest rate since the Great Depression. We use COVID-19 as an event study to compute the average treatment effect (ATE) of both voluntary and involuntary job exits on online educational content using augmented inverse propensity-weighted (AIPW) estimators. We found that following the pandemic, individuals who were involuntarily laid off increased their consumption of online educational content compared to people who were not laid off. We argue that this subgroup engages in online educational content in order to increase their human capital and have higher job prospects. Our significant findings are critical for policymakers when planning workforce initiatives during periods of mass layoffs.

Causal inference for environmental impact evaluation

A Novel Perspective for Carbon Offsetting through the Potential Outcomes Framework
Megan Ayers* Megan Ayers, Luke Sanford,

Forest carbon credits are purchased by entities seeking to offset negative environmental impacts by investing in conservation or restoration projects. Because offsets are generated by projects that avoid emissions or add sequestration, this practice inherently makes causal assumptions and relies on estimates of causal effects, though rarely formalized in practice. Offset calculations depend on emission estimates in counterfactual “baseline” scenarios where no crediting agreements are made. If observed emissions are reduced compared to these estimates, then projects are deemed “additional” and credits are awarded. Accounting for uncertainty and potential biases associated with baseline estimates is critical for the effectiveness of carbon offsetting as an environmental intervention, but current methodologies are inconsistent and often implicitly require unrealistic assumptions. In our work, we seek to close the gap between existing carbon offset methodologies and causal inference by providing: 1. A formalization of carbon offsetting practices within the potential outcomes framework, with accessibility for carbon offset stakeholders in mind; 2. Reformulations of existing baseline estimation practices within this framework; and 3. Conditions for unbiased and/or consistent estimation of the true amount of carbon offset under each baseline methodology. Finally, we assess over 30 complex offset protocols using the framework to provide guidance on when to value existing offsets.

Causal Inference in Networks

Randomized Experiment for Dyadic Data with Interference Yilin Li* Yilin Li, Lu Deng, Yong Wang, Wang Miao,

Estimating the global average treatment effect (total treatment effect) on a network could be considerably difficult in the presence of unknown network interference. We consider novel setting where the dyadic outcomes are available. Dyadic outcomes are common in many social network sources, such as forwarding a message or sharing a link. We first introduce the setting of network interference with unit-level treatment and dyadic outcomes, which is of particular interest in online experimentation. Then we manifest that the unbiased estimator for the global average treatment effect based on the unit-level outcomes does not exist in general. We provide subsequently unbiased estimators based on dyadic outcomes for randomized experiments. We show the possible variance bounds of our proposed estimators and provide an asymptotic conservative variance estimator. We illustrate the above phenomenon with a variety of numerical experiments. We utilize our method and discuss an application on the WeChat Channel.

Causal Inference in Networks

Inverse probability weighting-based bias correction methods for causal effects estimated under misspecified interference sets Laura Forastiere* Ariel Chao, Donna Spiegelman, Ashley Buchanan, Laura Forasteire,

Interference is often present in randomized or observational studies, where one participant's exposure to the intervention may affect another's outcome. To estimate causal effects, for each participant we must specify an interference set, that is, the set of those whose exposure may affect that participant's outcome. Interference sets are conventionally assumed to be correctly specified; however, they are prone to misspecification. For example, under the partial interference assumption, interference is assumed to be contained within well-separated clusters, yet social interactions may extend across them but are falsely assumed away. In this paper, we show that when interference sets are misspecified, causal effects estimated by an inverse probability weighting (IPW) estimator are biased. In HIV studies where social behaviors drive disease transmission, correcting causal effects for bias is crucial for accurate evaluation of interventions. We propose IPW-based bias-correction methods when a validation study containing data on the true interference sets is available, and extend these methods to the setting where multiple surrogates of the interference sets may be observed. We assessed finite sample properties of our methods in a simulation study, and applied the methods to the Botswana Combination Prevention Project, where clusters defined by two geographical boundaries were regarded as the surrogate interference sets, and phylogenetic data were used to define the true ones.

Causal Inference in Networks

Independent-Set Design of Experiments for Estimating Treatment and Spillover Effects under Network Interference Chencheng Cai* Chencheng Cai, Xu Zhang, Edoardo Airoldi,

Interference is ubiquitous when conducting causal experiments over networks. Except for certain network structures, causal inference on the network in the presence of interference is difficult due to the entanglement between the treatment assignments and the interference levels. In this article, we conduct causal inference under inter- ference on an observed, sparse but connected network, and we propose a novel design of experiments based on an independent set. Compared to conventional designs, the independent-set design focuses on an independent subset of data and controls their interference exposures through the assignments to the rest (auxiliary set). We provide a lower bound on the size of the independent set from a greedy algorithm , and justify the theoretical performance of estimators under the proposed design. Our approach is capable of estimating both spillover effects and treatment effects. We justify its superiority over conventional methods and illustrate the empirical performance through simulations.

Causal Inference in Networks

Harnessing the Scale of Spatial Confounding for Causal Inference with Areal Data Sophie Woodward* Sophie Woodward, Mauricio Tec,

Studies investigating the causal effects of continuous exposures on human health – such as air pollution, green space, or crime – often rely on observational and spatially-indexed areal data. A prevalent challenge is unmeasured spatial confounding, where an unobserved, spatially-varying variable affects both exposure and outcome, leading to biased causal estimates and invalid confidence intervals. Confounding might occur only at some spatial scales but not others. For instance, in studies on the health impacts of air pollution, local variations in healthcare may be less correlated with air pollution exposure after accounting for socioeconomic status and demographics, thus decreasing unmeasured confounding from healthcare at the local scale. Conversely, there are applications in which confounding is stronger locally but dissipates at coarser scales. We introduce a methodology to address spatial confounding bias restricted to certain spatial scales. By decomposing exposure into a component influenced by the unmeasured confounder and an independent component, based on assumptions about the scale of confounding, we can achieve causal identification of the exposure-response function by adjusting for the confounded component. This approach unifies existing literature where previous methods are special cases of our identifying functional. We also develop a sensitivity analysis framework that produces a sequence of estimators for different scale of confounding assumptions.

Causal Inference in Networks

Model-Based Inference and Experimental Design for Interference Using Partial Network Data Steven Wilkins-Reeves* Steven Wilkins-Reeves, Tyler McCormick, Arun Chandrasekhar, Shane Lubold,

The stable unit treatment value assumption states that the outcome of an individual is not affected by the treatment statuses of the individual's neighbors. In many common scenarios, ranging from economics to epidemiology, this assumption is not met. For instance, an individual's likelihood of being infected once given a vaccine likely depends on whether their close contacts received the vaccine. In many empirically relevant situations, full network data (required to adjust for these spillover effects) is too costly or logistically infeasible to collect. Partially or indirectly observed network data (e.g., subsamples, Aggregated Relational Data (ARD), egocentric sampling, or respondent-driven sampling) reduce the logistical and financial burden of collecting network data, but the statistical properties of treatment effect adjustments from these design strategies were, until now, largely unknown. In this paper, we present a framework for the estimation and inference of treatment effect adjustments using partial network data. Further, we demonstrate how to use partial network data to inform randomization in experimental settings to reduce the variance of the treatment effect estimate. In addition to our theoretical results, we evaluate this approach using simulated experiments on observed graphs, as well as an application to information diffusion.

Causal Inference in Networks

Engineering Social Groups: How Organizational Group Assignment Affects the Social Fabric Johanna Einsiedler* Johanna Einsiedler, Andreas Bjerre-Nielsen, Nikolaj Arpe Harmon, Jolien Cremers, David Dreyer Lassen,

There has been a long-standing discussion about the relevance of social networks in higher education for student outcomes. To date, a large body of literature exists that points towards a substantial role of peer effects in academic achievement social behavior and occupational choice. However, to leverage peer effects in the design of effective policies to, e.g., reduce achievement gaps, detailed knowledge about the underlying causes and mechanisms of peer effects is necessary. Studying these has traditionally been hard, as most data available to researchers lacks information about the actual social interactions among students. Existing work has had to rely on either organizational assignment, e.g. to dorms or rooms or self-reported connections for determining peer groups.
In our study, we investigate how group assignment effects both in-person and online social interactions. We leverage a novel dataset that combines high-resolution measures of social interactions (including physical proximity, calling/texting, and Facebook friendship data) with organizational assignment information. To measure the immediate and persistent effects of group assignment, we exploit an initial random placement of the participating students into groups upon admission to college. We further study the interaction of those group assignment effects with other known drivers of tie formation such as homophily and triadic closure using a network formation model based on subgraphs.

Causal inference with continuous exposures

A new causal estimand for continuous exposures Anand Hemmady* Anand Hemmady, Marco Carone, Andrea Rotnitzky,

To study the causal effect of a continuous point-exposure on an outcome of interest, investigators often contrast mean counterfactual outcomes under different exposure patterns. Most commonly, exposure patterns under which all participants are assigned the same exposure value are considered. While it results in intuitively interpretable estimands, such an approach has been critiqued for various reasons, including that it can lead to exposure patterns that would be impossible in the real world, which violates positivity. This has led to the development of alternative frameworks. In this work, motivated by recent work on modified treatment policies and stochastic interventions, we define a novel exposure pattern by specifying the cumulative odds ratio between the new and factual exposure patterns. The proposed exposure pattern has a desirable interpretation while avoiding violations of positivity. We derive an identification for the causal parameter resulting from the proposed exposure pattern and propose a debiased machine learning approach for inference. We also consider the identification and inference of the parameter under commonly encountered complications, including censoring and two-phase sampling, and use the resulting methodology to analyze data from COVID-19 vaccine efficacy trials.

Causal inference with high-dimensional covariates

Causal Inference with High-dimensional Discrete Covariates Zhenghao Zeng* Zhenghao Zeng, Edward Kennedy, Sivaraman Balakrishnan, Yanjun Han,

When estimating causal effects from observational studies, covariate adjustment is often required to deconfound the non-causal relationship between exposure and outcome (i.e., association). For modern datasets featuring an increasing number of covariates, researchers may have access to discrete covariates (with potentially a large number of categories), where commonly assumed structures such as smoothness fail to hold and the behavior of popular regression, weighting and doubly robust estimators has not been well-understood. In this work, we study estimation of the causal effect in a model where the covariates required for confounding adjustment are discrete but high-dimensional, meaning the number of categories diverges. Specifically, we study the theoretical properties of commonly used estimators mentioned above and provide sufficient and necessary conditions for them to be consistent. We also consider additional structures that can be exploited, namely effect homogeneity and prior knowledge on covariate distribution, and propose new estimators that enjoy faster convergence rate and achieve consistency in a broader regime. The results are illustrated empirically via simulation studies. Importantly, we also derive minimax lower bound of the average treatment effects, which characterizes the fundamental difficulty of causal effects estimation in high-dimensional discrete setting.

Difference in Differences

Survival after hospitalization: Constructing counterfactual time-to-event outcomes for difference-in-differences studies Laura Hatfield* Laura Hatfield, Bret Zeldow,

Improving the health care and outcomes of hospitalized patients is the goal of many health policy interventions. To estimate the effects of these interventions in non-randomized settings, many researchers turn to difference-in-differences designs, which contrast observed outcomes with counterfactual outcomes imputed under a “parallel trends” assumption. Inspired by difference-in-differences designs, we propose a novel strategy to impute a counterfactual survival curve by leveraging the evolution of a comparison group’s survival curve. Our strategy can identify survival time estimands like the difference in median survival. We contrast our novel approach with estimands based on event risks/rates, which can be identified with simple additive and multiplicative versions of parallel trends. In an analysis of survival data for hospitalized patients, we compare feasibility, plausibility, and usefulness of the methods. We show that although estimands based on event risks/rates are simple to identify and estimate, they can obscure important patterns in the presence of censoring. By contrast, our proposed strategy identifies the whole survival curve and estimands based on any summary of it and therefore can correctly account for censoring using simple Kaplan-Meier-based estimation methods.

Difference in Differences

Beyond parallel trends: unifying difference-in-differences and synthetic controls Denis Agniel* Denis Agniel, Max Rubinstein, Jessie Coe, Maria DeYoreo,

We propose a new method for estimating causal effects in longitudinal/panel data settings that we call stable bias difference-in-differences. Our approach unifies two alternative approaches in these settings: ignorability estimators (e.g., synthetic controls) and difference-in-differences (DiD) estimators. We propose a new identifying assumption — a stable bias assumption — which generalizes the conditional parallel trends assumption in DiD, leading to the proposed stable bias DiD framework. This change gives stable bias DiD estimators the flexibility of ignorability estimators while maintaining the robustness to unobserved confounding of DiD. We also show how ignorability and DiD estimators are special cases of stable bias DiD. We then propose influence-function based estimators of the observed data functional that identifies the average treatment effect on the treated, allowing the use of double/debiased machine learning for estimation. We also show how stable bias DiD easily extends to include clustered treatment assignment and staggered adoption settings, and we discuss how the framework can facilitate estimation of other treatment effects beyond the average treatment effect on the treated. Finally, we provide simulations which show that stable bias DiD outperforms ignorability and DiD estimators when their identifying assumptions are not met, while being competitive with these special cases when their identifying assumptions are met.

Generalizability/Transportability

Generalizing the intention-to-treat effect of an active control against placebo from historical placebo-controlled trials to an active-controlled trial: A case study of the efficacy of daily oral TDF/FTC in the HPTN 084 study Qijia He* Qijia He,

In many clinical settings, an active-controlled trial design is often used to compare an experimental medicine to an active control. One prominent example is a recent phase 3 efficacy trial, HIV Prevention Trials Network Study 084 (HPTN 084), comparing long-acting cabotegravir, a new HIV pre-exposure prophylaxis (PrEP) agent, to the FDA-approved daily oral tenofovir disoproxil fumarate plus emtricitabine (TDF/FTC) in a population of heterosexual women in 7 African countries. One key complication of interpreting study results in an active-controlled trial is that the placebo arm is not present and the efficacy of the active control compared to the placebo can only be inferred by leveraging other data sources. In this article, we study statistical inference for the intention-to-treat (ITT) effect of the active control using relevant historical placebo-controlled trials data under the potential outcomes (PO) framework. We highlight the role of adherence and unmeasured confounding, discuss in detail identification assumptions and two modes of inference (point versus partial identification), propose estimators under identification assumptions permitting point identification, and lay out sensitivity analyses needed to relax identification assumptions. We applied our framework to estimating the intention-to-treat effect of daily oral TDF/FTC versus placebo in HPTN 084 using data from an earlier placebo-controlled trial of daily oral TDF/FTC (Partners PrEP).

Generalizability/Transportability

Transporting average causal effects with positivity violations Paul Zivich* Paul Zivich, Jessie Edwards, Bonnie Shook-Sa, Eric Lofgren, Justin Lessler, Stephen Cole,

Transportability methods can be used to estimate causal effects from a biased sample of the target population. Transportability relies on a positivity assumption, such that all relevant covariate patterns in the target population also occur in the secondary population from which the sample was selected. Strict eligibility criteria, particularly in the context of randomized trials, can lead to violations of this assumption. Common methods to address nonpositivity are to restrict the target population, restrict the adjustment set, or extrapolate from a statistical model. Instead of these approaches, which all have concerning limitations, we propose a synthesis of statistical (e.g., g-methods) and mathematical (e.g., mechanistic) models. Briefly, a statistical model is fit for the regions of the parameter space where positivity holds, and a mathematical model is used to fill-in the nonpositive regions. For estimation, we propose two novel augmented inverse probability weighting estimators; one based on a marginal structural model, and the other based on the conditional average causal effect. The proposed methods are applied to estimate the effect of antiretroviral therapy on CD4 cell count among women with HIV. The synthesis approach addresses positivity violations when transporting and may be extended for other applications in causal inference more generally.

Generalizability/Transportability

Efficient combination of observational and experimental datasets with structural information about outcome mean functions Harrison Li* Harrison Li,

We consider a setting where a researcher running a randomized experiment additionally has access to possibly biased or confounded observational data, and seeks to leverage prior structural knowledge about the outcome mean functions in the two datasets to improve estimation of average treatment effects in a target population defined by the subjects in either or both of the datasets. For example, a known transformation of these mean functions may be assumed to be transportable or differ by a low-dimensional parametric component between the two datasets. Through an explicit characterization of the tangent space of influence functions of regular asymptotically linear estimators, we provide a succinct yet general framework for deriving semiparametric efficiency bounds under such assumptions, providing novel insight into the types of structural information that can and cannot underpin efficiency gains. We then use this characterization to construct efficient estimators by computationally projecting known estimators' influence functions onto this tangent space, thereby avoiding the need to explicitly derive the efficiency bounds or estimate nuisance variance components. We apply this procedure to develop novel efficient estimators when the observational dataset is subject to a parametric confounding bias or outcome-mediated selection bias. The finite sample performance of our estimators is demonstrated through simulations and data examples used in prior work.

Generalizability/Transportability

Disparate effect of missing mediators on the transportability of causal effects Vishwali Mhasawade* Vishwali Mhasawade, Vishwali Mhasawade, Rumi Chunara,

Transported mediation effects provide an avenue to understand how upstream interventions, as opposed to proximal interventions, such as individual behaviors, would work differently when applied to different populations. However, when mediators are missing in the populations where the effect is to be transported, these estimates could be biased, with the conditional average treatment effect being insignificant for the subgroup with missing mediator data. We study this issue of missing mediators, motivated by challenges in public health where the mediators are commonly missing, not at random. We propose a sensitivity analysis framework that quantifies the impact of missing mediator data on transported mediation effects. This framework enables us to identify the settings under which the conditional transported mediation effect is rendered insignificant for the subgroup with missing mediator data. Specifically, we provide the bounds on the transported mediation effect as a function of missingness. We then apply the sensitivity framework to longitudinal data from the Moving to Opportunity Study, a large-scale housing voucher experiment, to transport effect estimates of voucher receipt, an upstream intervention on living location, in childhood on subsequent risk of mental health or substance use disorder mediated through parental health across sites. Our findings contribute to a tangible understanding of how much missing data can be withstood for unbiased effect estimates.

Generalizability/Transportability

Completeness results for identification in selected data models for data fusion Jaron Lee*
Jaron Lee, AmirEmad Ghassami, Ilya Shpitser,

Data fusion has become an increasingly important aspect of causal inference, as combining different datasets can lead to improved identification and estimation capabilities. We consider data fusion problems where selection into different datasets is potentially systematic. We propose an explicit modeling of such systematic selection, yielding a hierarchy of selection problems analogous to the missing data hierarchy: selection completely at random (SCAR), selection at random (SAR), and selection not at random (SNAR). Importantly, it enables us to consider the task of identification and estimation in a coherent “full data distribution” that incorporates the selection mechanisms. We propose a graphical representation of the constraints in this full data model using which we provide a sound and complete identification algorithm for identification of causal effects from combinations of observational and experimental domains where the selection process and variables in these domains are potentially confounded in complicated ways.

Graphical models

Unpacked graphs for causal inference Sonia Markes* Sonia Markes, Elie Wolfe,

Physicists have studied latent variable models for almost a century to understand quantum phenomena. They have recently made significant progress, showing how causal compatibility can be explicitly formulated for some causal structures involving latent variables, such as in Bell experiments in quantum physics or the instrumental variable scenario known in causal inference. Our research showcases how these advancements can be utilized for causal inference. We have developed unpacked graphs to depict all the causal links between counterfactual variables. By utilizing the independence relations encoded in these unpacked graphs, we can obtain narrower partial identifiability bounds on the causal estimand. In this presentation, we will demonstrate how this method of using unpacked graphs to obtain bounds for a few example causal structures.

Heterogeneous Treatment Effects

Flexibly Estimating and Interpreting Heterogeneous Treatment Effects of Laparoscopic Surgery for Cholecystitis Patients Luke Keele* Luke Keele, Matteo Bonvini, Zhenghao Zeng, Edward Kennedy,

Recent methodological work has developed a meta-learner framework for flexible estimation of conditional causal effects. In this framework, nonparametric estimation methods can be used to avoid bias from model misspecification while preserving statistical efficiency. In addition, researchers can flexibly and effectively explore whether treatment effects vary with a large number of possible effect modifiers. However, these methods have certain limitations. For example, conducting inference can be challenging if black-box models are used. Further, interpreting and visualizing the effect estimates can be difficult when there are multi-valued effect modifiers. In this paper, we develop new methods that allow for interpretable results and inference from the meta-learner framework for heterogeneous treatment effects estimation. We also demonstrate methods that allow for an exploratory analysis to identify possible effect modifiers. We apply our methods to a large database for the use of laparoscopic surgery in treating cholecystitis. We also conduct a series of simulation studies to understand the relative performance of the methods we develop. Our study provides key guidelines for the interpretation of conditional causal effects from the meta-learner framework.

Heterogeneous Treatment Effects

Qini Curves for Multi-Armed Treatment Rules Erik Sverdrup* Erik Sverdrup, Han Wu, Susan Athey, Stefan Wager,

Qini curves have emerged as an attractive and popular approach for evaluating the benefit of data-driven targeting rules for treatment allocation. We propose a generalization of the Qini curve to multiple costly treatment arms, that quantifies the value of optimally selecting among both units and treatment arms at different budget levels. We develop an efficient algorithm for computing these curves and propose bootstrap-based confidence intervals that are exact in large samples for any point on the curve. These confidence intervals can be used to conduct hypothesis tests comparing the value of treatment targeting using an optimal combination of arms with using just a subset of arms, or with a non-targeting assignment rule ignoring covariates, at different budget levels. We demonstrate the statistical performance in a simulation experiment and an application to treatment targeting for election turnout.

Heterogeneous Treatment Effects

Active Feature Acquisition in Precision Medicine Michael Valancius* Michael Valancius, Michael Kosorok, Junier Oliva,

A fundamental goal in precision medicine is to learn individualized treatment rules (ITRs) that use rich biomarker data to tailor treatment recommendations. When treatment effects are heterogenous with respect to demographic, genetic, or clinical data, these personalized actions can lead to improved health outcomes. While this motivates learning ITRs that are functions of rich and complex data, the collection of covariate information in routine practice is often dynamic and bears a cost (financial, time, inconvenience, etc.). Therefore, paradigms basing ITRs on static sets of features can have limited real-world applicability. In this work, we consider tailoring the feature acquisition process to an individual. We provide conditions under which currently acquired features are informative for deciding whether (and which) additional covariates would be beneficial to collect. Furthermore, we show that the objective function representing the expected health outcomes under a given a feature acquisition policy and an ITR can be cast as a weighted classification problem, enabling the usage of standard machine learning methods to learn these two policies. Simulation results demonstrate the empirical improvement of this approach compared to alternative approaches that do not tailor the feature acquisition process based on the observed individual characteristics.

Heterogeneous Treatment Effects

LOOL: A Simple and Precise Estimator of Heterogeneous Treatment Effects Duy Pham* Duy Pham, Adam Sales,

Given the ever-present and growing demand for personalized treatment and intervention in healthcare, education, policy, and many other domains, there have been significant developments in methodology to effectively and efficiently measure heterogeneous treatment effects over the last decade. We propose the Leave-One-Out Learner (LOOL) – a new meta-learner approach for the estimation of conditional average treatment effects (CATEs). Given a correctly specified experiment design, we can first obtain unbiased estimates of the individual treatment effects (ITEs) – without requiring any additional assumptions – by applying the Leave-One-Out Potential Outcomes (LOOP) Estimator given by Wu and Gagnon-Bartsch (2018). By regressing these estimates on the covariates in the treatment and control group, we further obtain the expected CATE functions under each condition. Like the X-Learner given by Künzel et al (2019), each observation's CATE estimate is the sum of the corresponding values of these two conditional functions – weighted by the propensity score. However, ITEs from LOOP are unbiased, and regressing them on the covariates combats the relatively higher variance. Thus, LOOL can potentially obtain more accurate estimates. Compared to existing meta-learners, LOOL's performance was highly competitive in a wide range of initial simulations.

Heterogeneous Treatment Effects

Combining T-learning and DR-learning: a framework for oracle-efficient estimation of causal contrasts Lars van der Laan* Lars van der Laan, Marco Carone, Alex Luedtke,

We introduce efficient plug-in (EP) learning, a novel framework for the estimation of heterogeneous causal contrasts, such as the conditional average treatment effect and conditional relative risk. The EP-learning framework enjoys the same oracle-efficiency as Neyman-orthogonal learning strategies, such as DR-learning and R-learning, while addressing some of their primary drawbacks, including that (i) their practical applicability can be hindered by loss function non-convexity; and (ii) they may suffer from poor performance and instability due to inverse probability weighting and pseudo-outcomes that violate bounds. To avoid these drawbacks, EP-learner constructs an efficient plug-in estimator of the population risk function for the causal contrast, thereby inheriting the stability and robustness properties of plug-in estimation strategies like T-learning. Under reasonable conditions, EP-learners based on empirical risk minimization are oracle-efficient, exhibiting asymptotic equivalence to the minimizer of an oracle-efficient one-step debiased estimator of the population risk function. In simulation experiments, we illustrate that EP-learners of the conditional average treatment effect and conditional relative risk outperform state-of-the-art competitors, including T-learner, R-learner, and DR-learner.

Heterogeneous Treatment Effects

A Practical Minimax Approach to Causal Inference under Limited Overlap Yuanzhe Ma*
Yuanzhe Ma, Yian Huang, Hongseok Namkoong,

A central challenge in observational analysis is that the effective sample size may be prohibitively small when there is little overlap between treated and control populations. Data sparsity becomes more pronounced in modern operational problems that involve high-dimensional covariates. Existing observational methods, which truncate data with extreme propensity scores, are only valid in the large sample limit and silently fail in practical instances with limited effective sample size. In this work, we propose a new inferential framework that provides always-valid uncertainty quantification, which provides a sensitivity analysis framework against unforeseen data sparsity. Our work builds on the theory of minimax estimation for linear functionals that can generate an always-valid confidence interval. We operationalize the above minimax approach by decomposing the data into overlap and non-overlap regions. We use standard asymptotic inference tools (e.g., AIPW) on the region of overlap, and only apply the conservative minimax approach described above on the non-overlap region while also utilizing information from the overlap region. Through simulated and real data, we demonstrate our method ensures robustness against unforeseen data sparsity by quantifying the bias induced by standard truncation techniques.

Heterogeneous Treatment Effects

Developing and validating a treatment recommendation score with an application to the heterogeneous impact of transesophageal echocardiography on clinical outcomes post coronary artery bypass graft surgery Charlotte Talham* Charlotte Talham, Emily MacKay, Qijia He, Bo Zhang,

While past observational research has shown transesophageal echocardiography (TEE) to improve patient outcomes post coronary artery bypass graft (CABG) surgery, resource constraints in clinical settings limit the use of TEE in practice. Our aim is thus to adopt a precision medicine approach to better assign TEE to patients who could most benefit from its use. Two practical challenges emerge. First, to be useful to practitioners, a summary of the heterogeneous treatment effect is needed to inform decisions about how to allocate limited resources. Second, validation of the proposed precision-TEE approach is critical to promoting its integration into medical practice. We present a framework that aims to address both challenges. We first perform an exploratory analysis on a random subset of data in which we take a machine learning approach to develop a treatment recommendation score (TRS) through solving a series of weighted classification problems, each subject to varying resource constraints. Next, we use the remainder of the data to validate the findings in a matched cohort study in which participants who received TEE are compared to those who did not receive TEE, but who were assigned the same TRS. Our matching procedure balances a large number of clinical and demographic covariates. The proposed framework is illustrated using synthetic data modeled after the Society of Thoracic Surgeons (STS) national registry. This method is also being used to analyze the STS registry data.

Heterogeneous Treatment Effects

Synergizing Experiments: Designing Personalized Marketing Interventions through Incrementality Representation Learning Ta-Wei Huang* Ta-Wei Huang, Eva Ascarza, Ayelet Israeli,

In the pursuit of personalization, firms aim to tailor interventions to match individual customer preferences. Traditional methods typically consist of two stages: first, a set of predefined interventions are tested and evaluated across various customer segments; then, the most effective intervention for each segment is assigned. Although this approach provides some benefits of personalization, they often fall short in achieving complete customization, constrained by the capacity to test only a few interventions within a set number of segments. Furthermore, these methods do not provide insights for designing new interventions or for targeting different segments, thereby restricting the scope and potential of data-driven personalization.

This research introduces a novel approach enabling companies to design personalized interventions more efficiently by capitalizing on historical experimental data. We develop a flexible causal machine learning framework, termed incrementality representation learning, which estimates the conditional average treatment effect (CATE) based on intervention characteristics and customer covariates and extracts low-dimensional representations of these features to capture the heterogeneity in treatment effects. This approach allows firms to leverage past experiments to identify the most effective type of intervention for any specific

Heterogeneous Treatment Effects

Inference on Variable Importance Measures for Heterogeneous Treatment Effects Pawel Morzywolek* Pawel Morzywolek, Alex Luedtke,

Recent years have seen a growing interest in quantifying treatment effect heterogeneity, which is vital for supporting individualized decision-making. Though black-box machine learning approaches might optimally predict treatment effect heterogeneity, in high-risk domains such as medicine, decision makers often hesitate to rely on decision support systems without understanding the underlying rationale behind the recommendations. Hence, it is crucial to offer insights into which variables best predict individualized treatment effects. Motivated by these considerations, we present model-agnostic variable importance measures for heterogeneous treatment effects. We provide efficient estimators of these measures together with corresponding confidence intervals, and introduce a Wald-type test to assess the null hypothesis of no importance. Our approach builds on recent developments in semiparametric theory for pathwise differentiable function-valued parameters, and is valid even when flexible black-box algorithms are employed to quantify treatment effect heterogeneity. We demonstrate the applicability of our methodology in the context of infectious disease prevention strategies.

Heterogeneous Treatment Effects

Uniform Inference for Local Conditional Quantile Treatment Effect Curve with High-Dimensional Covariates Jing Tao* Jing Tao,

In this study, we investigate heterogeneous local quantile treatment effects for observational data with high-dimensional covariates without relying on the strong ignorability assumption. Using a binary instrumental variable, the parameters of interest are in a population subgroup (compliers) through a two-stage regression model. We develop Lasso estimation with a non-convex and non-smooth objective function to estimate the parameters of interest. We propose a de-sparsifying estimator for pointwise and uniform inference for quantile treatment effects. Moreover, we obtain uniform strong approximations of the local quantile treatment process by conditionally pivotal and Gaussian processes. Based on these strong approximations, we develop bootstrap resampling methods to construct uniform confidence bands for the heterogeneous/conditional local quantile treatment effects given high-dimensional covariates. Finally, we evaluate performance through simulation studies.

Heterogeneous Treatment Effects

Fully Latent Principal Stratification With Measurement Models Adam Sales* Adam Sales, Sooyong Lee, Hyeon-Ah Kang, Tiffany Whittaker,

There is wide agreement on the importance of implementation data from randomized effectiveness studies in behavioral science; however, there are few methods available to incorporate these data into causal models, especially when they are multivariate or longitudinal, and interest is in low-dimensional summaries. We introduce a framework for studying how treatment effects vary between subjects who implement an intervention differently, combining principal stratification with latent variable measurement models; since principal strata are latent in both treatment arms, we call it “fully-latent principal stratification” or FLPS. We describe FLPS models including item-response-theory measurement, show that they are feasible in a simulation study, and illustrate them in an analysis of hint usage from a randomized study of computerized mathematics tutors.

Heterogeneous Treatment Effects

GEEPERs: Principal Stratification using Principal Scores and Stacked Estimating Equations Adam Sales* Adam Sales, Kirk Vanacore, Erin Ottmar,

Principal stratification is a framework for making sense of causal effects conditioned on variables that themselves may have been affected by treatment. For instance, one component of an educational computer application is the availability of “bottom-out” hints that provide the answer. In evaluating a recent experimental evaluation against alternative programs without bottom-out hints, researchers may be interested in estimating separate average treatment effects for students who, if given the opportunity, would request bottom-out hints frequently, and for students who would not. Most principal stratification estimators rely on strong structural or modeling assumptions, and many require advanced statistical training to fit and check. In this paper, we introduce a new M-estimation principal effect estimator for one-way noncompliance based on a binary indicator. Estimates may be computed using conventional regressions (though the standard errors require a specialized sandwich formula) and do not rely on distributional assumptions. We present a simulation study that shows that the novel method is more robust than popular alternatives and illustrate the method in an analysis of data on bottom-out hint requests.

Heterogeneous Treatment Effects

Causal inference with continuous treatments: a tale of two estimands Oliver Hines* Oliver Hines, Stijn Vansteelandt, Karla Diaz-Ordaz,

In economics and medicine one is often interested in the main effect of a continuous treatment (dose, price, duration) on an outcome. Dose-response curve modelling is a popular approach, but the uniform interventions considered may be unrealistic for some subpopulations. Worse still, estimators may be poorly supported by the data due to the extrapolation needed to predict expected outcomes under unrealistic treatments. This talk focuses on two emerging alternatives. Average Derivative Effects (ADEs) indicate the mean direction and magnitude of small changes to the treatment around the values observed in the data. Least Squares Estimands (LSEs) represent the “coefficient” when the outcome is projected (in a nonparametric sense) on to a partially linear model. In our work we showed that LSEs are optimally weighted ADEs and both belong to a common class that also contains the average treatment effect of a binary treatment on an outcome. This poster gives some motivation and intuition behind these estimands and results.

Heterogeneous Treatment Effects

Treatment Effect Heterogeneity with Spatio-temporal Data Lingxiao Zhou* Lingxiao Zhou, Kosuke Imai, Jason Lyall, Georgia Papadogeorgou,

Understanding the causal effects of processes involving spatial and temporal dimensions has become increasingly important in various fields, including ecology and epidemiology. However, the classical methods used to estimate treatment effect heterogeneity are not directly applicable in this unique setting. We consider the case where treatment and outcome are spatio-temporal point patterns with arbitrary dependence in space and time, and effect modifiers are spatial, temporal, or spatio-temporal variables. This work introduces causal estimands for treatment effect heterogeneity in this challenging setting and proposes a two-stage estimation strategy, that draws from literature on causal inference with spatio-temporal data and estimation methods for treatment effect heterogeneity for independent and identically distributed data. We propose a least squares estimator based on weighted outcomes with stabilized weights and we use martingale theory to show that the proposed estimator is asymptotically normal as the number of time periods increases. We illustrate the method by studying treatment effect heterogeneity for the relationship between American airstrikes and insurgent violence in Iraq.

Instrumental Variables

Long-Term Causal Inference with Imperfect Surrogates using Many Weak Experiments, Proxies, and Cross-Fold Moments Aurelien Bibaut* Nathan Kallus, Aurelien Bibaut, Simon Ejdemyr, Michael Zhao,

Inferring causal effects on long-term outcomes using short-term surrogates is crucial to rapid innovation. However, even when treatments are randomized and surrogates fully mediate their effect on outcomes, it’s possible that we get the direction of causal effects wrong due to confounding between surrogates and outcomes — a situation famously known as the surrogate paradox. The availability of many historical experiments offer the opportunity to instrument for the surrogate and bypass this confounding. However, even as the number of experiments grows, two-stage least squares has non-vanishing bias if each experiment has a bounded size, and this bias is exacerbated when most experiments barely move metrics, as occurs in practice. We show how to eliminate this bias using cross-fold procedures, JIVE being one example, and construct valid confidence intervals for the long-term effect in new experiments where long-term outcome has not yet been observed. Our methodology further allows to proxy for effects not perfectly mediated by the surrogates, allowing us to handle both confounding and effect leakage as violations of standard statistical surrogacy conditions.

Instrumental Variables

Manipulating a Continuous Instrumental Variable: Algorithm, Partial Identification Bounds, and Inference under Randomization and Biased Randomization Assumptions Min Haeng Cho* Min Haeng Cho, Zhe Chen, Bo Zhang,

An instrumental variable (IV) can be thought of as a random nudge towards accepting a treatment. With a continuous IV, Baiocchi et al. (2010) strengthened the original IV using non-bipartite matching and proposed a valid test for the effect ratio estimand, an analog of the sample average treatment effect (SATE) among compliers. Their key insight is to shift focus from the entire study cohort to a possibly smaller cohort amenable to being paired with a larger separation in the IV dose, inducing a higher compliance rate. Three elements change as one switches from one design to the other. First, the study cohort changes. In this article, we show it can be avoided using a template matching algorithm. Second, the compliance rate changes. Third, the latent complier subgroup changes as a person's principal stratum status in a matched design is defined with respect to the two IV doses within each pair. In this article, we study partial identification bounds for the SATE for the entire matched cohort. Unlike the effect ratio, the SATE estimand does not depend on who is matched to whom in the design, although a strengthened-IV design may narrow its partial identification bounds. We derive valid statistical inference for the partial identification bounds under a randomization assumption and an IV-dose-dependent, biased randomization scheme in a matched-pair design, with applications to a study of the effect of neonatal intensive care units on the mortality rate of premature babies.

Instrumental Variables

Instrumental Variable Estimation in Compositional Regression for Time-Use Surveys in Health and Long-Term Care Andrej Srakar* Andrej Srakar,

Time use surveys are used in many areas of economics, including economics of health and long-term care. If analyzed in a regression context, time use survey data suffer from the problem of spurious correlation noted in early works of Aitchison (1986). This problem leads to a need for compositional regression perspective on a geometric simplex. We develop an instrumental variable compositional regression model, building on two strands of literature with applications for health economics and economics of long-term care. We extend Florens and Van Bellegem (2015) functional instrumental variables model to compositional data setting where either or both independent and dependent variables are of compositional nature. We show there exist two ways of deriving compositional IV's, one using isometric log-ratio transform and Chesher et al. (2013)'s IV model of multiple discrete choice; and another deriving from the recent literature on compositional functional data in Bayes spaces (Machalova et al., 2021). We show that estimation leads to an ill-posed inverse problem with a data-dependent operator and we use and extend the notion of instrument strength to compositional setting. We establish appropriate functional CLT's and study the finite sample performance in a Monte Carlo simulation setting. Our application studies relationship between long term care for older people and paid work, using recent time use survey from Survey of Health, Ageing and Retirement in Europe (SHARE).

Instrumental Variables

The Instrumental Variable Model with a Binary Outcome and Categorical Instrument and Treatment Yilin Song* Yilin Song, Kwun Chuen Gary Chan, Thomas Richardson,

Instrumental variable models are central to the inference of causal effects in many settings, including Mendelian randomization and clinical trials with non-compliance. Richardson and Robins (2014) studied the instrumental variable model with binary exposure (X) and binary outcome (Y) with an instrument (Z) that takes k states where k□2. In our work, we consider the instrumental variable model allowing X to be categorical with p□2 states, while Y still being binary and Z taking k states. We assume that the instrument is randomized and that there is no direct effect of Z on Y so that Y(x,z) = Y(x). We first provide a simple characterization of the set of joint distributions of the potential outcomes p(Y(x=1),..., Y(x=p)) compatible with a given observed probability distribution p(X, Y|Z). We then characterize the resulting constraints on the margins p(Y(x=1)), ..., p(Y(x=p)). We find that, in contrast to the case where p=2, these margins are not variation-independent when p>2. We discuss the implications for partial identification of average causal effect contrasts such as E[Y(x=i) – Y(x=j)].

Instrumental Variables

Experimental Design with a Binary Instrument Under a Partially Linear Model Tim Morrison* Tim Morrison, Minh Nguyen, Michael Baiocchi, Art Owen,

We study the question of how best to assign an encouragement in a randomized encouragement study. In our setting, units arrive with covariates, receive a “nudge” toward treatment or control, acquire one of those statuses in a way that need not align with the nudge, and finally have a response observed. The nudge can be seen as a binary instrument that affects the response only via the treatment status.

Our interest is in choosing how to assign the nudge as a function of covariates to best estimate the local average treatment effect (LATE). We assume a partially linear model, wherein the baseline model is non-parametric and the treatment term is linear in the covariates. Under this model, we outline a two-stage procedure to consistently estimate the LATE. This leads to a finite sample approximation of the variance and thus a design criterion to minimize. This criterion is convex, allowing for constraints that might arise for budgetary or ethical reasons. We prove conditions under which our solution asymptotically recovers the lowest variance.

Our motivating example comes from triage in emergency departments (EDs), in which nurses sort patients to either the intensive care unit (ICU) or less serious care. A natural question is the causal effect of being sent to the ICU, but this is confounded with the potential outcomes. We describe a randomized encouragement study to assess this question and apply our method on a semi-synthetic dataset of ED patients to estimate the LATE.

LLM and Causality

A New Frontier at the Intersection of Causality and LLMs Emre Kiciman* Emre Kiciman, Robert Ness, Amit Sharma, Chenhao Tan,

Correct causal reasoning requires domain knowledge beyond observed data. Consequently, the first step to correctly frame and answer cause-and-effect questions in medicine, science, law, and engineering requires working closely with domain experts and capturing their (human) understanding of system dynamics and mechanisms. This is a labor-intensive practice, limited by expert availability, and a significant bottleneck to widespread application of causal methods.

In this talk, we will delve into the causal capabilities of large language models (LLMs), discussing recent studies and benchmarks of their ability to retrieve and apply causal knowledge, as well as the limitations of their causal reasoning capabilities. Most notably, LLMs present the first instance of general-purpose assistance for constructing causal arguments, including generating causal graphical models and identifying contextual information from natural language. This promises to reduce the necessary human effort and error in end-to-end causal inference and reasoning, broadening their practical usage. Ultimately, by capturing common sense and domain knowledge, we believe LLMs are a catalyst for a new frontier facilitating translation between real world scenarios and causal questions, and formal and data-driven methods to answer them.

Machine Learning and Causal Inference

Economic Burden of Breast Cancer in Denmark: Estimation and Evaluation of Casual Methods Emily Johnson* Emily Johnson, Angela Chang, Liza Sopina,

The economic burden of breast cancer has long been a topic of interest within the literature, and in Nordic countries the existence of the extensive administrative data systems allows for causal analysis of this topic. Prior studies have applied matching to administrative data estimate the effects of disease burden on income. However, none of these studies have empirically evaluated the dependence of these results on the methods used.
This paper applies machine learning to evaluate the validity of different causal models in measuring the cost of breast cancer in Denmark. Data are sourced from the national administrative registries which capture the Danish population from 2000-2018. Matching methods evaluated include propensity score matching, coarsened exact matching, inverse probability weighting, and Bayesian additive regression trees, and matching parameters tested include age, socioeconomic status, education, family size, and baseline household income. Models are compared using synthetic validation, which can approximate cross validation for predictive estimation. These metrics provide more information about model fit than evaluations of matching performance, which can assess the quality of a match but not the bias of a measured causal effect. By applying synthetic validation to a topic of interest in health economics literature, this paper demonstrates a mechanism to evaluate the validity of causal inference results where empirical assessment of bias is often lacking.

Machine Learning and Causal Inference

Optimal nonparametric estimation of constrained functional parameter Razieh Nabi* Razieh Nabi, Nima Hejazi, Mark van der Laan, David Benkeser,

Statistical machine learning algorithms, integral to sectors like hiring, finance, and healthcare, risk reinforcing societal biases based on gender, race, religion, among others. To combat this, it's vital to design models adhering to fairness norms. This involves embedding fairness constraints such as 'equal opportunity', ensuring uniform true positive rates across groups, and 'counterfactual fairness', assessing outcomes in varied hypothetical scenarios. This study doesn't favor a specific fairness criterion but proposes a general framework for deriving optimal prediction functions under various constraints. It conceptualizes the learning problem as estimating a constrained functional parameter within a comprehensive statistical model, using a Lagrange-type penalty. This enables representing a fair prediction function in relation to an unfair counterpart, plus other parameters, for integration with standard learning frameworks. Key contributions of our work include a flexible framework for solving constrained optimization problems, closed-form solutions for specific fairness constraints, and an algorithm-neutral approach to fair learning. This framework's applicability extends beyond algorithmic fairness to other constrained learning contexts like Neyman-Pearson classification, churn reduction, adversarial learning, and reinforcement learning, demonstrating broad utility and impact.

Machine Learning and Causal Inference

Post-Episodic Reinforcement Learning Inference Ruohan Zhan* Ruohan Zhan, Vasilis Syrgkanis,

We consider estimation and inference with data collected from episodic reinforcement learning (RL) algorithms; i.e. adaptive experimentation algorithms that at each period (aka episode) interact multiple times in a sequential manner with a single treated unit. Our goal is to be able to evaluate counterfactual adaptive policies after data collection and to estimate structural parameters such as dynamic treatment effects, which can be used for credit assignment (e.g. what was the effect of the first period action on the final outcome). Such parameters of interest can be framed as solutions to moment equations, but not minimizers of a population loss function, leading to Z-estimation approaches in the case of static data. However, such estimators fail to be asymptotically normal in the case of adaptive data collection. We propose a re-weighted Z-estimation approach with carefully designed adaptive weights to stabilize the episode-varying estimation variance, which results from the nonstationary policy that typical episodic RL algorithms invoke. We identify proper weighting schemes to restore the consistency and asymptotic normality of the re-weighted Z-estimators for target parameters, which allows for hypothesis testing and constructing uniform confidence regions for target parameters of interest. Primary applications include dynamic treatment effect estimation and dynamic off-policy evaluation.

Machine Learning and Causal Inference

Causal Importance of Features in Machine Learning Bo Liu* Bo Liu, Fan Li,

Measuring the importance of predictors or features on the outcome variable is crucial in supervised learning. Existing measures of feature importance rely solely on the black-box models and are often sensitive to model misspecification, because they require extrapolation to predict the outcome for values far outside of the training set. The degree of necessary extrapolation depends on the overlap between the remaining predictors at two levels of a particular predictor, which is also a key concept in causal inference. Common approaches in causal inference to reduce sensitivity to outcome model specification include alternative methods such as propensity score weighting and double robust estimators, and alternative estimands such as the average treatment effect on population with sufficient overlap. We borrow these ideas and point out the connections between several commonly used feature importance measures in machine learning literature and causal estimands. We then propose several new robust measures of feature importance for black-box supervised learning models.

Machine Learning and Causal Inference

Estimating Heterogenous Treatment Effects for Survival Data with Doubly Doubly Robust Estimator Guanghui Pan* Guanghui Pan,

In this paper, we introduce a doubly doubly robust estimator for the average and heterogeneous treatment effect for left-truncated-right-censored (LTRC) survival data. In causal inference for survival functions in LTRC survival data, two missing data issues are noteworthy: one is the missing data of counterfactuals for causal inference, and the other is the missing data due to truncation and censoring. Based on previous research on non-parametric deep learning estimation in survival analysis, this paper proposes an algorithm to obtain an efficient estimate of the average and heterogeneous causal effect. We simulate the data and compare our methods with the marginal hazard ratio estimation, the naive plug-in estimation, and the doubly robust causal with Cox Proportional Hazard estimation and illustrate the advantages and disadvantages of the model application.

Machine Learning and Causal Inference

Evaluating the impact of popular films on public interest in and consumption of plant-based food Anna Thomas* Anna Thomas, Maya Mathur, Jessica Hope,

Due to the harms of factory farming to public health, the environment, and animal welfare, many organizations, including the UN and the Intergovernmental Panel on Climate Change, have called for a shift to a plant-based diet. Interventions such as educational campaigns, alternative protein, and policy changes may serve to accelerate this shift. Here we focus on evaluating educational campaigns in the form of documentary films, in order to provide a recommendation to the organizations we work with on whether further promotion of specific existing films (e.g. via screenings or advertisements) or creation of new films are likely to be effective. Several of these films received press on anecdotal reports of behavior change.

We have access to national-level time series on exposures, various outcomes, and time-varying covariates. Using this observational data, we aim to estimate short-term effects of point interventions on film exposure on the outcomes. Results from our pre-registered approach using a doubly robust, propensity score-adjusted regression model, a special case of the structural nested mean model, suggest contemporaneous and lagged effects of some films on Google search volume for plant-based food. However, we do not observe an effect on meat demand. In ongoing work, we are studying the impact on sales of plant-based and animal-based food via the National Consumer Panel, as well as assessing sensitivity of our findings to unmeasured confounding and other assumptions.

Machine Learning and Causal Inference

Robust inference for the treatment effect variance in experiments using machine learning
Alejandro Sanchez Becerra* Alejandro Sanchez Becerra,

Experimenters often collect baseline data to study heterogeneity. I propose the first valid confidence intervals for the VCATE, the treatment effect variance explained by observables. Conventional approaches yield incorrect coverage when the VCATE is zero. As a result, practitioners could be prone to detect heterogeneity even when none exists. The reason why coverage worsens at the boundary is that all efficient estimators have a locally-degenerate influence function and may not be asymptotically normal. I solve the problem for a broad class of multistep estimators with a predictive first stage. My confidence intervals account for higher-order terms in the limiting distribution and are fast to compute. I also find new connections between the VCATE and the problem of deciding whom to treat. The gains of targeting treatment are (sharply) bounded by half the square root of the VCATE. Finally, I document excellent performance in simulation and reanalyze an experiment from Malawi.

Machine Learning and Causal Inference

Comprehensive Causal Machine Learning Jana Mareckova* Jana Mareckova, Michael Lechner,

Uncovering causal effect heterogeneity across various levels of granularity is invaluable for decision-makers. Comprehensive causal estimation approaches enable the use of a single machine learning method to estimate effects at all granularity levels, making them attractive for applied studies due to their computational tractability and unified empirical framework.

In this paper, we compare three comprehensive approaches: double machine learning (DML), generalized random forest (GRF), and modified causal forest (MCF). DML provides a generic framework for estimating (conditional) average causal effects using ML methods. MCF estimates conditional average causal effects and utilizes weighted representations for higher aggregation, while GRF employes doubly robust estimators for aggregates. The paper provides theoretical results for MCF including weight-based inference for causal effects and their asymptotic normality.

A simulation study examines these approaches, considering selection into treatment, effect heterogeneity and other properties of the data generating process, and provides valuable finite sample insights. Metrics include bias, standard deviation, root MSE and coverage probabilities. DML excels in estimating effects at higher aggregation levels. GRF exhibits higher bias with moderate to high treatment selection, while MCF shows robust performance across treatment selection scenarios, exhibiting lower bias than GRF and better or similar coverage probabilities.

Machine Learning and Causal Inference

Super Ensemble Learning Using the Highly-Adaptive-Lasso: Imaging Data in Causal Inference Zeyi Wang* Zeyi Wang, Wenxin Zhang, Mark van der Laan,

Imaging-based models have great potential to enable more powerful and efficient causal analysis. However, it is challenging to incorporate high-dimensional models into causal inference under different scientific contexts. In this paper, we present a novel minimum loss estimation framework with meta-learning and the Highly Adaptive Lasso (HAL). For a true functional parameter defined as the minimizer of the expectation of a loss function, we consider ensemble estimators that are compositions of a cadlag function and a data adaptive coordinate-transformation. Meta-HAL minimum loss estimator is defined as the cadlag function that minimizes the cross-validated empirical risk of the ensemble estimator, over all cadlag functions with a uniform bound on the sectional variation norm. The average of the resulted ensemble estimators across folds is called meta-HAL super-learner. We show that under regularity conditions, the meta-HAL super-learner converges to the true function at a rate n^{−2/3} up till log n-factor in the excess risk, and by choosing the sectional variation norm large enough the target feature of the meta-HAL super-learner is an asymptotically linear estimator for the target feature of the true function. This leads to effective dimension reduction for which we provide theoretical guarantees, simulation evidence with concrete examples including average treatment effects, and real-world evidence of natural indirect effects of functional brain images in pain studies.

Machine Learning and Causal Inference

Ranking multiple treatment effects using double-machine-learning estimators Apoorva Lal*
Apoorva Lal, Winston Chou, Jordan Schafer,

We study the properties of prominent semi-parametric ('Double Machine Learning') estimators with many treatments in settings where decision-makers want to learn a ranking of causal effects for downstream decisions-rules using a series of numerical examples. We show that the conditional-variance-Weighted Average Treatment Effects (WATEs) yielded by the partially-linear model (PLM) may not have the same ranking as underlying Average Treatment Effects (ATEs), and therefore may result in suboptimal decisions. In contrast, we find that Augmented Inverse-Propensity-Weighting (AIPW) and Automatic Debiased Machine-Learning (Auto-DML) estimators that perform explicit rather than implicit weighting rank underlying ATEs correctly, and therefore should be preferred in applications that require treatment ranking.

Machine Learning and Causal Inference

Local Longitudinal Modified Treatment Policies Herbert Susmann* Herbert Susmann, Iván Díaz,

Longitudinal Modified Treatment Policies (LMTPs) provide a framework for defining a broad class of causal target parameters. We propose Local LMTPs, a generalization of LMTPs to settings where the target parameter is conditioned on the assigned treatment. Such parameters have wide scientific relevance, with well-known parameters such as the Average Treatment Effect on the Treated (ATT) falling within the class. We provide a formal causal identification result that expresses the Local LMTP parameter in terms of sequential regressions, and derive the efficient influence function of the parameter which defines its semi-parametric efficiency bound. Efficient semi-parametric inference of Local LMTP parameters requires estimating the ratios of functions of complex conditional probabilities (or densities). We propose an estimator for Local LMTP parameters that directly estimates these required ratios via empirical loss minimization, drawing on the theory of Riesz representers. The estimator is implemented using a combination of ensemble machine learning algorithms and deep neural networks, and evaluated via simulation studies.

Machine Learning and Causal Inference

Hidden among subgroups: Detecting critical treatment effect bias in observational studies
Piersilvio De Bartolomeis* Piersilvio De Bartolomeis, Javier Abad, Konstantin Donhauser, Fanny Yang,

Randomized trials are considered the gold standard for making informed decisions in medicine. However, they are often not representative of the patient population in clinical practice. Observational studies, on the other hand, cover a broader patient population but are prone to various biases. Thus, before using observational data for any downstream task, it is crucial to emph{benchmark} its treatment effect estimates against a randomized trial.
We propose a novel strategy to benchmark observational studies on a subgroup level. First, we design a statistical test for the null hypothesis that the treatment effects — conditioned on a subset of relevant features — differ up to some tolerance value. Our test allows us to estimate an asymptotically valid lower bound on the maximum bias strength for any subgroup. We validate our lower bound in a real-world setting and show that it leads to conclusions that align with established medical knowledge.

Machine Learning and Causal Inference

Evaluating immune correlates of protection in vaccine efficacy trials with stochastic-interventional causal effects Nima Hejazi* Nima Hejazi, Peter Gilbert,

In vaccine efficacy clinical trials randomizing participants to active v control conditions and following individuals until the occurrence of an infectious disease outcome of interest, evaluating the efficacy of a candidate vaccine through immune markers specified a priori is complicated by factors including (1) the relative dearth of causal inference approaches tailored to quantitative mediators and (2) the challenge of correcting for outcome-dependent (e.g., case-cohort) sampling used to measure such markers. We present a two-pronged solution, using (1) modified treatment policies to formulate interventions on immune markers whose identifiability can be rigorously probed and (2) inverse probability of sampling weights to obtain population-level inference. We outline non-/semi-parametric inference strategies that yield asymptotically efficient estimators of our proposed causal estimands and that incorporate machine learning. Our effect definitions measure the causal effect of perturbing an immune marker’s observed value in the active condition, resulting in an interpretable causal dose-response analysis informing on potential vaccine protection mechanisms and aiding in identification of surrogate endpoints. We demonstrate their utility by highlighting applications to modeling how modifications to a vaccine tested in a phase 3 trial would be expected to alter vaccine efficacy across four distinct trials of the Coronavirus Prevention Network and HIV Vaccine Trials Network.

Machine Learning and Causal Inference

Evaluating the effect of lifting COVID-19 eviction moratoria on drug overdose death rates: an implementation of targeted learning through a modified treatment policies approach

Ariadne Rivera Aguirre* Ariadne Rivera Aguirre, Giselle Routhier, Ivan Diaz, Kelly Doran, Magdalena Cerda,

During the first two years of the COVID-19 pandemic, the US witnessed 207,315 avoidable drug overdose deaths. Higher eviction rates have been associated with higher drug overdose deaths. In light of the economic turmoil caused by the pandemic, and the subsequent increased risk of eviction, 44 states implemented temporary eviction moratoria. Many of these protective measures expired between April 2020-December 2021, leaving vulnerable households at risk of eviction. The effects of these housing interventions on drug overdoses remain unexplored.
We evaluated if the expiration of state eviction moratoria was associated with higher county overdose death rates from April 2020–December 2021 in the US. We obtained mortality data from NCHS and eviction moratoria dates from the COVID-19 US State Policy Database. We addressed positivity violations and time-varying confounding in this volatile period, using a modified treatment policies algorithm by implementing a longitudinal targeted minimum loss-based estimation with SuperLearner to estimate the ATE and the effect of a 1-month delay lifting the eviction moratoria. Initial results suggest that the ATE of lifting eviction moratoria on monthly overdose deaths is 0.2 additional deaths per 100,000 people (95%CI: 0.13, 0.27). A 1-month delay lifting moratoria could have decreased monthly deaths by 0.19 per 100,000 people (95%CI: -0.33, -0.04). These findings imply that even temporary housing measures can help curb the overdose epidemic.

Matching

Re-evaluating the impact of hormone replacement therapy on heart disease using match-adaptive randomization inference Sam Pimentel* Sam Pimentel, Ruoqi Yu,

Matching is an appealing way to design observational studies because it mimics the data structure produced by stratified randomized trials, pairing treated individuals with similar controls. After matching, inference is often conducted using methods tailored for stratified randomized trials in which treatments are permuted within matched pairs. However, in observational studies, matched pairs are not predetermined before treatment; instead, they are constructed based on observed treatment status. This introduces a challenge as the permutation distributions used in standard inference methods do not account for the possibility that permuting treatments might lead to a different selection of matched pairs (Z-dependence). To address this issue, we propose a novel and computationally efficient algorithm that characterizes and enables sampling from the correct conditional distribution of treatment after an optimal propensity score matching, accounting for Z-dependence. We show how this new procedure, called match-adaptive randomization inference, corrects for an anticonservative result in a well-known observational study investigating the impact of hormone replacement theory (HRT) on coronary heart disease and corroborates experimental findings about heterogeneous effects of HRT across different ages of initiation in women. Supported by the National Science Foundation under Grant No. 2142146.

Matching

Cardinality Matching with Multiple Versions of Treatment Lauren Liao* Lauren Liao, Amanda Ngo, Emily Wang, Rana Chehab, Yeyi Zhu, Samuel Pimentel,

Even when a study's primary research question focuses on the effect of a binary treatment, it is common for some variation to arise in how treatment is delivered. For example, in a study comparing a medication to standard of care, different brands of that medication may be given to subjects. Effects of individual versions of treatment (e.g. one particular brand) are often of interest as secondary analyses, in addition to the effect of the overall composite treatment. However, a study that finds a balanced comparison between treated and control subjects for the overall treatment may not provide balanced comparisons for the individual versions of treatment. We propose a study design that creates a single matched comparison with balance both for the overall treatment comparison and for comparisons using individual versions of the treatment. We leverage cardinality matching to impose constraints on both overall and subgroup balance, and use a tuning parameter to encode a researcher's relative preference between the two types of balance. We demonstrate this method using a large cohort study of pregnant subjects to evaluate the impact of categorizing subjects into stage 1 or stage 2 hypertension based on serial blood pressure measurements on adverse perinatal outcomes. Our analysis targets simultaneously the overall balance between stage 1 and stage 2 hypertension and subgroup balance, where stage 2 hypertension divides into individual versions with or without medication.

Matching

Random network in time series studies with bipartite interference Zhaoyan Song* Zhaoyan Song, Georgia Papadogeorgou,

In bipartite causal inference with interference there are two separate sets of units: those that receive the treatment, termed interventional units, and those on which the outcome is measured, termed outcome units. Which interventional units' treatment can drive which outcome units' outcomes is often depicted in a bipartite graph. We study bipartite causal inference with interference from data across time and a changing network. We establish unconfoundedness of the exposure received by the outcome units based on unconfoundedness assumptions on the interventional units' treatment assignment and the random graph, hence respecting the bipartite structure of the problem. We define causal effects at the level of the outcome units. By harvesting the time component of our setting, we propose a causal effect identification strategy that is outcome unit-specific, and it requires controlling only for temporal trends and time-varying confounders. Our identification results hold for binary, continuous, and multivariate exposures. In the case of a binary exposure, we propose three matching algorithms to estimate the causal effect based on matching exposed to unexposed time periods, and we show that the bias of the resulting estimators is bounded. We illustrate our approach with an extensive simulation study and an application on the effect of the presence of wildfire smoke on transportation by bicycle.

Mediation

Powerful Partial Conjunction Hypothesis Testing via Conditioning Biyonka Liang* Biyonka Liang, Lu Zhang, Lucas Janson,

The testing of causal hypotheses, such as in mediation analysis and settings involving evidence factors, is often formulated as a Partial Conjunction Hypothesis (PCH) test, which combines information across a set of base hypotheses to determine whether some subset is non-null. However, standard methods for testing a PCH can be highly conservative. In this paper, we introduce the conditional PCH (cPCH) test, a new framework for testing a single PCH that directly corrects the conservativeness of standard approaches by conditioning on certain order statistics of the base p-values. Under distributional assumptions commonly encountered in PCH testing, the cPCH test produces uniform null p-values. Through simulations, we demonstrate that the cPCH test uniformly outperforms standard single PCH tests, maintains Type I error control even under model misspecification, and, in certain settings, can also be used to outperform state-of-the-art multiple testing approaches for causal mediation analysis.

Mediation

Causal Mediation Analysis with Ultra-high Dimensional Potential Confounders for the Study on Geriatric Depression and Alzheimer's Disease Yuexia Zhang* Yuexia Zhang, Annie Qu, Yubai Yuan, Qi Xu, Fei Xue, Kecheng Wei,

Depression and Alzheimer's Disease (AD) are both prevalent diseases in older adults. Using the data sets from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, we explore whether geriatric depression has a significant average treatment effect on AD and whether the effect is mediated by some important mediators. To estimate these causal effects consistently, we control for ultra-high dimensional potential confounders, including DNA methylation levels. We propose a new ball correlation-based screening method for confounder selection in mediation analysis. To achieve robustness against model misspecification, we utilize a robust mediation analysis framework. Simulation studies show that the proposed method has good finite-sample performance in terms of confounder and mediator selection, effect estimation, and inference. In the real data analysis, we find that geriatric depression has a significantly positive causal effect on AD. We also propose new prevention and treatment strategies for geriatric depression and AD through changing the selected confounders and mediators.

Mediation

Multiply Robust Causal Mediation Analysis with Continuous Treatments AmirEmad Ghassami* Numair Sani, AmirEmad Ghassami, Yizhen Xu, Ilya Shpitser,

In many applications, researchers are interested in the direct and indirect causal effects of a treatment or exposure on an outcome of interest. Mediation analysis offers a rigorous framework for identifying and estimating these causal effects. For binary treatments, efficient estimators for the direct and indirect effects are presented in Tchetgen Tchetgen and Shpitser (2012) based on the influence function of the parameter of interest. These estimators possess desirable properties, such as multiple-robustness and asymptotic normality, while allowing for slower than root-n rates of convergence for the nuisance parameters. However, in settings involving continuous treatments, these influence function-based estimators are not readily applicable without making strong parametric assumptions. In this work, utilizing a kernel-smoothing approach, we propose an estimator suitable for settings with continuous treatments inspired by the influence function-based estimator of Tchetgen Tchetgen and Shpitser (2012). Our proposed approach employs cross-fitting, relaxing the smoothness requirements on the nuisance functions, and allowing them to be estimated at slower rates than the target parameter. Additionally, similar to influence function-based estimators, our proposed estimator is multiply robust and asymptotically normal, making it applicable for inference in settings where a parametric model cannot be assumed.

Mediation

Unpack the mechanisms of treatment effects in AB tests Can Cui* Can Cui,

Treatment effects measured in AB tests provide an objective and valuable way to evaluate performance of product changes or engineering improvements, and to guide launch decisions. While standard estimates of treatment effect is important, we sometimes run in to the scenario where treatment effects are counter intuitive, or difficult to explain. In the space of streaming experiences, we run thousands of tests to validate and evaluate various engineering changes. Better understanding of mechanisms of measured treatment effects can help tremendously to further improve engineering interventions and success rate of future tests. As most tests involve complex changes (e.g. involving changes in multiple parts of the system) and lead to changes in customer experience in multiple dimensions, we leverage various methods to unpack the underlying mechanisms of the treatment effects, including instrumental variables method, decomposition analysis, and mediation analysis. We will discuss the pro and con of these approaches, especially highlighting the identification assumptions in different types of AB tests, and showcase real life examples of learning that help engineers tweak future interventions and come up with newer and better tests.

Missing data

Making predictions from missing data, without the imputation Tanvi Shinkre* Tanvi Shinkre, Chad Hazlett,

When using covariate data (X) to model some outcome (Y), missing values in X are often problematic. If observations with missingness cannot be dropped without causing problematic biases, investigators often employ imputation approaches. These methods model the relationship among covariates so as to fill in missing values with predictions. In some settings, this is done repeatedly to incorporate uncertainty in the prediction of missing values (multiple imputation). However, kernel-based regression methods suggest an alternative approach, which can utilize the available X information to model Y without imputation or dropping observations. Methods such as kernel ridge regression model Y by relying only on the similarity of each observation to other observations. Such similarity can be measured between any units i and j over just the covariates present in both observations. With an appropriate scaling we discuss, this allows Y to be modeled using all the available information in X, despite missingness, without imputation. We show that in the case of the "linear kernel", this produces exactly the same predictions for Y_i as an unregularized linear regression among units with no missingness, while also producing informative predictions for Y_i for units with some missingness in X. Other kernels, such as the Gaussian kernel, can similarly be adapted to employ only the co-present covariates for each pairwise similarity measure with appropriate rescaling.

Multilevel Causal Inference

Estimating exposure-response curves in multi-level data: causal pooling of cluster-specific curves Jenny Lee* Jenny Lee, Fabrizia Mealli, Francesca Dominici, Rachel Nethery,

When studying the health impacts of fine particulate matter (PM_{2.5}) at a national level, one of the key interests of policy-makers may be how different regions contribute to exposure-response function (ERF). Substantial variability on health outcomes across regions after controlling for measured confounders may be due to differences in the unmeasured regional-level variables such as cultural difference, dietary habits, and prevalence of comorbidity. In this paper, we propose pooled-ERF, a method for population-level causal ERF estimation with multi-level data, which allows for adjustment of unmeasured cluster-level confounders, enables estimation of the population-level ERF from cluster-level ERFs even when the raw data for each cluster is inaccessible (i.e. only have cluster level data), and yields an interpretable ERF that elucidates each cluster’s contribution. We compare commonly used ERF estimators across different confounding scenarios and compare performance of pooled-ERF via simulation. We apply pooled-ERF to estimate the average causal ERF between long-term PM_{2.5}exposure and all-cause mortality among 68.5 million Medicare enrollees 2000-2016 based on the regional level ERFs across four regions in the U.S. (northeast, west, south, midwest). We demonstrate the accuracy and interpretable nature of the results using the tools and visualization techniques enabled by the proposed method, and we compare results to those of other causal inference methods.

Propensity Scores

Reporting practices and guidelines for machine learning to estimate propensity scores
Walter Leite* Walter Leite, Huibin Zhang, Zachary Collier, Kamal Chawala, Lingchen Kong, YongSeok Lee, Jia Quan, Olushola Soyoye,

Non-parametric machine learning (ML) methods, such as generalized boosting and random forests, have been extensively used in propensity score analysis (PSA). Their key advantage over parametric methods, such as logistic regression, consists of automatically detecting complex relationships between a larger number of covariates and the treatment assignment mechanism. Also, ML can prevent multicollinearity problems that arise from using strongly correlated sets of covariates. However, detailed guidelines on reporting the use of ML for propensity score estimation in academic papers do not exist. This study developed guidelines for ML reporting in PSA based on a systematic review of over 150 peer-reviewed papers, dissertations, and theses published from 1983 to 2023 across social sciences, health sciences, and education. The guidelines are aligned with best practices in open science and research reproducibility and are organized by the following six steps of PSA: 1) data preparation, 2) propensity score estimation, 3) propensity score method implementation, 4) covariate balance evaluation, 5) treatment effect estimation, and 6) sensitivity analysis. The systematic review shows that few published papers provide enough details about their use of ML for PSA to allow replication of analyses.

Propensity Scores

Antihypertensive target trial emulation in electronic health records to inform drug repurposing alternatives for dementia prevention Marie-Laure Charpignon* Marie Laure Charpignon, Bella Vakulenko-Lagun, Colin Magdamo, Bowen Su, Sudeshna Das, Anthony Philippakis, Munther Dahleh, Deborah Blacker, Ioanna Tzoulaki, Mark Albers,

Alzheimer’s disease, the most common type of dementia, affects 6.7 million Americans and costs $345B annually. Since disease-modifying therapies are limited, repurposing FDA-approved drugs may offer an alternative, expedited path to preventing dementia. Hypertension is a major risk factor for dementia onset. However, prior observational studies contrasting antihypertensive drug classes (Angiotensin Converting Enzyme inhibitors: ACEI, Angiotensin Receptor Blockers: ARB, and Calcium Channel Blockers: CCB), provided mixed results. We hypothesize that ACEI have an off-target pathogenic mechanism. To test this assumption, we emulate a target trial comparing patients initiating ACEI vs ARB using electronic health records from the US Research Patient Data Registry. We perform intention-to-treat analyses among patients aged 50+, applying Inverse Propensity score of Treatment Weighting to balance the two treatment arms and accounting for competing risk of death. In a cause-specific Cox Proportional Hazards (PH) model, the hazard of dementia onset was lower in ARB vs ACEI initiators (HR=0.72 [95% CI: 0.68-0.77]). Findings were robust to outcome model structures (i.e., Cox PH vs nonparametric) and generalized to patients with no hypertension diagnosis at initiation. Our trial emulation suggests that ARB initiation may reduce risk of dementia onset. Future work will evaluate differential effects by brain penetrance and the mediating role of blood pressure control in dementia prevention.

Randomized Studies

Test-Negative Designs with Various Reasons for Testing: Statistical Bias and Solution.
Mengxin Yu* Mengxin ¥u,

Test-negative designs are widely used for post-market evaluation of vaccine effectiveness. Different from classical test-negative designs where only healthcare-seekers with symptoms are included, recent test-negative designs have involved individuals with various reasons for testing, especially in an outbreak setting. While including these data can increase sample size and hence improve precision, concerns have been raised about whether they will introduce bias into the current framework of test-negative designs, thereby demanding a formal statistical examination of this modified design. In this article, using statistical derivations, causal graphs, and numerical simulations, we show that the standard odds ratio estimator may be biased if various reasons for testing are not accounted for. To eliminate this bias, we identify three categories of reasons for testing, including symptoms, disease-unrelated reasons, and case contact tracing, and characterize associated statistical properties and estimands. Based on our characterization, we propose stratified estimators that can incorporate multiple reasons for testing to achieve consistent estimation and improve precision by maximizing the use of data. The performance of our proposed method is demonstrated through simulation studies.

Randomized Studies

Identifying localized treatment effects in high-dimensional outcome spaces Yujin Jeong*
Yujin Jeong, Ramesh Johari, Emily Fox,

Based on technological advances in sensing modalities, randomized trials with primary outcomes represented as high-dimensional vectors have become increasingly prevalent. For example, these outcomes could be week-long time-series data from wearable devices or neuro-signal graph data derived from magnetic resonance imaging. This paper focuses on randomized treatment studies with such high-dimensional outcomes characterized by localized treatment effects, where interventions may influence a small number of dimensions, e.g., small temporal windows or specific clustered brain regions. Conventional practices, such as using fixed low-dimensional summaries for outcomes, result in significantly reduced power for detecting treatment effect signals. To address this limitation, we propose a procedure that involves subset selection followed by inference. Specifically, given a set of direction vectors, we identify the subset that captures treatment effects and subsequently conduct inference on these selected directions. Via theoretical analysis as well as simulations, we demonstrate that our method asymptotically selects the correct subset and increases statistical power.

Randomized Studies

Experimental Design For One-sided Matching Marketplaces Nian Si* Nian Si, Chenran Weng, Xiao Lei,

One-sided matching markets, prevalent in scenarios where users are matched with other users, are evident in environments like video game platforms and anonymous social networks. Here, participants are matched for interactions such as games or social exchanges. Experimentation (A/B tests) in these markets is challenging due to the interdependence of users' metrics on their counterparts' treatment assignments. In this paper, we build a stochastic market model and develop its mean field limit to analyze such experimental dynamics. Our focus is on two randomization strategies: user and match randomization. We demonstrate that, under Markovian conditions and homogeneous users behavior, match randomization provides unbiased estimations but can lead to significant biases when these conditions are not met. Conversely, user randomization shows greater resilience to model inaccuracies. We further propose an associated linear regression estimator that can halve the bias compared to a naive estimator.

Randomized Studies

Tackling Interference Induced by Data Training Loops in A/B Tests: A Weighted Training Approach NIAN SI* NIAN SI,

In modern recommendation systems, the standard pipeline involves training machine learning models on historical data to predict user behaviors and improve recommendations continuously. However, these data training loops can introduce interference in A/B tests, where data generated by control and treatment algorithms, potentially with different distributions, are combined. To address these challenges, we introduce a novel approach called weighted training. This approach entails training a model to predict the probability of each data point appearing in either the treatment or control data and subsequently applying weighted losses during model training. We demonstrate that this approach achieves the least variance among all estimators without causing shifts in the training distributions. Through simulation studies, we demonstrate the lower bias and variance of our approach compared to other methods.

Randomized Studies

Design-based inference for paired cluster-randomized experiments Charlotte Mann*
Charlotte Mann, Adam Sales, Johann Gaganon-Bartsch,

Paired cluster-randomized experiments (pCRTs) are common in education and medicine because there is a natural clustering of students/patients within classrooms/hospitals and within schools/cities. Additionally, paired randomization can help balance baseline covariates to improve experimental precision. Although paired cluster-randomization is very common, there is surprisingly no obvious way to analyze this randomization design if an individual-level (rather than cluster-level) treatment effect is of interest. For example, the most basic and common estimator, the difference in mean outcomes, is biased in this setting. Variance estimation is also complicated due to the dependency created through pairing clusters in pCRTs. In order to provide guidance for practitioners analyzing pCTS, first, we review point estimators and associated variance estimators for an individual-level sample average treatment effect for pCRTs, unifying the notation. We propose a design-based point estimator and variance estimator, and show how this estimator relates and, in fact, provides a unifying framework for previous estimators in the literature. Through analysis based on this framework and extensive simulation studies, we illustrate the trade-offs between the reviewed point and variance estimators in practice.

Randomized Studies

Safety first: Design-informed inference with `propertee` Ben Hansen* Ben Hansen, Joshua Errickson, Xinhe Wang, Joshua Wasserman,

In studies with social or medical data, units of analysis may be divisible into “clusters”, for instance students nested within classrooms or patient observations nested within patients. Despite its importance to inference, cluster composition can be hazy, even arbitrary.

In randomized controlled trials (RCTs), as well as those observational studies that model themselves on RCTs (Rubin, 2008; Hernan & Robins, 2016), investigators specify concrete units of assignment in the process of articulating the study’s design (Rosenbaum, 2010). In most analyses of such evaluation studies, the appropriate clusters are simply the units of assignment; in all such studies, appropriate clustering keeps assignment units intact.

Our “propertee” R package aims to facilitate several seemingly straightforward impact estimation tasks that can difficult to execute safely. These include making use of predictions from a model fitted to external or partially external samples, and bringing in design-based inverse probability weights. Its method includes separate elicitation of study design information, identifying the assignment units as a matter of course. This makes it easy and safe to, for example, produce Hajek estimates and associated standard errors for RCTs with large or small blocks — even in the presence of grouped assignment to treatment, irregular repeated measurements, or subgroup-level estimation with subgroups seen only under treatment in some blocks and only under control in others.

Randomized Studies

Design-based analysis of Hajek estimators with covariate adjustment in stratified and clustered RCTs Xinhe Wang* Xinhe Wang, Ben Hansen,

When subjects are naturally or artificially aggregated in clusters in randomized controlled trials (RCTs), clustered RCTs are conducted. To reduce the impact of imbalance of baseline covariates across different treatment groups, stratification or regression adjustment is advocated. Our previous work has shown that the difference of Hajek estimators is a robust estimator of the average treatment effect (ATE) in stratified and clustered RCTs when clusters are not matched on sizes within strata. In this work, we focus on incorporating covariance adjustment to the difference of Hajek estimators. Our approach involves fitting a prior regression model based on outcomes and covariates, followed by applying inverse probability weights to estimate the ATE. We compare the proposed method with existing ones for estimating the treatment effect in stratified and clustered RCTs such as Schochet et al (2022, JASA), and outline the scenarios where the Hajek estimator with covariance adjustment is more robust compared to other estimation methods. We verify the consistency of this estimator and propose a sandwich-type standard error estimator for it.

Randomized Studies

Criteria-Based Randomization: A Flexible and Transparent Restricted Randomization Framework for Better Experimental Design Maggie Wang* Maggie Wang, René Kizilcec, Michael Baiocchi,

Randomized experiments are considered the gold standard for estimating causal effects. However, out of the set of possible randomized assignments, some may be likely to produce poor effect estimates and misleading conclusions. Restricted randomization is an experimental design strategy that filters out undesirable treatment assignments, but its application has primarily been limited to ensuring covariate balance in two-arm studies where the target estimand is the average treatment effect. Other experimental settings with different design desiderata and target effect estimands could also stand to benefit from a restricted randomization approach. We introduce Criteria-based Randomization (CBR), a transparent and flexible framework for restricted randomization that filters out undesirable treatment assignments based on analyst-specified, domain-informed design criteria. Notably, CBR permits diverse design criteria beyond those that relate to balance, while preserving valid randomization inference. In CBR, the acceptable treatment assignments are locked in ex ante and pre-registered in the trial protocol, thus safeguarding against p-hacking and promoting reproducibility. Through illustrative simulation studies motivated by education and behavioral health interventions, we demonstrate how CBR can be used to improve effect estimates compared to benchmark experimental designs in three settings: multi-arm experiments, group formation experiments, and experiments with interference.

Randomized Studies

Forward selection and post-selection inference in factorial designs Lei Shi* Lei Shi, Peng Ding, Jingshen Wang,

Ever since the seminal work of R. A. Fisher and F. Yates, factorial designs have been an important experimental tool to simultaneously estimate the effects of multiple treatment factors. In factorial designs, the number of treatment combinations grows exponentially with the number of treatment factors, which motivates the forward selection strategy based on the sparsity, hierarchy, and heredity principles for factorial effects. Although this strategy is intuitive and has been widely used in practice, its rigorous statistical theory has not been formally established. To fill this gap, we establish design-based theory for forward factor selection in factorial designs based on the potential outcome framework. We not only prove a consistency property for the factor selection procedure but also discuss statistical inference after factor selection. In particular, with selection consistency, we quantify the advantages of forward selection based on asymptotic efficiency gain in estimating factorial effects. With inconsistent selection in higher-order interactions, we propose two strategies and investigate their impact on subsequent inference. Our formulation differs from the existing literature on variable selection and post-selection inference because our theory is based solely on the physical randomization of the factorial design and does not rely on a correctly specified outcome model.

Regression Discontinuity

The impact of the long-term indoor thermal environment exposure from China’s Huai River Policy Ruiji Sun* Ruiji Sun,

This paper explores the adaptive thermal comfort theory, which posits that individuals in hotter climates, or those exposed to warm indoor environments over extended periods, may develop a higher comfort temperature compared to those in colder climates. The adaptive comfort model demonstrates a correlation between outdoor air temperatures and indoor comfort temperatures. However, correlation is not causation. Do outdoor conditions or prolonged exposure to specific indoor temperatures causally affect our comfort levels? This study seeks to address this question by examining the winter heating policy near China’s Qin Mountain and Huai River, a demarcation line between two major climatic zones. In the northern zone, the lowest monthly mean outdoor air temperature falls below 0°C, while in the southern zone, it remains above 0°C. The heating policy enables buildings in the north to access affordable heating systems from centralized plants, typically maintaining indoor temperatures above 15°C. Conversely, buildings in the south, lacking such systems, have indoor temperatures only marginally above the outdoor temperature, near 0°C. We employ a regression discontinuity design based on monthly outdoor air temperatures to estimate the local causal impact of long-term exposure to warm indoor environments on occupants’ comfort temperature at the boundary of Qin Mountain and Huai River. The findings of this study significantly influence building standards and codes.

Regression Discontinuity

An application of regression discontinuity design for evaluating the impact of program features in a computer-based learning platform Kirk Vanacore* Kirk Vanacore, Adam Sales, Ben Hansen,

Regression Discontinuity Design (RDD) is commonly employed in economics, public policy, and education research, but it is underutilized in human-computer interaction research. Yet, features in computer programs are often good candidates for RDD, because many decisions in these programs are made based on a cut point. In the area of computer-based learning platforms (CBLPs), some examples include the administering of rewards, prescribing usage recommendations, and determining whether students have mastered a knowledge component. These mechanisms, which determine how students experience learning programs, provide opportunities for understanding the impacts of these features on users' behaviors and outcomes.

In the current study, we evaluated the impact of game-based failure on students' persistence behavior in a gamified CBLP using a regression discontinuity design. We found that game-based failure increases the likelihood of students engaging in productive persistence as they play an online gamified algebra program. This finding suggests that gamification features in learning programs help sidestep the negative aspects of failure and leverage those failure experiences for learning. This work also illustrates the usefulness of regression discontinuity designs in evaluating the impact of features in online learning games to provide insight into causal mechanisms through which program features influence students' learning processes.

Sensitivity Analysis

Introducing the specificity score: a measure of causality beyond P value Wang Miao* Wang Miao,

There is considerable debate about P value in scientific research and its use is banished in several prestigious journals in recent years. Particularly in observational studies where confounding arises, P value as a measure of statistical significance fails to capture the causal association of scientific interest. In this talk, I will introduce a specificity score for testing the existence of causal effects in the presence of unmeasured confounding. The specificity score measures how extreme the observed association is when compared to the confounding bias. A large specificity score means the observed association cannot be explained away by confounding and is thus evidence of causality. Under certain conditions, the specificity test has controlled type I error and power approaching unity for testing the null hypothesis of no causal effect. This approach only entails certain rough information on the broadness of the causal associations, but does not require the availability of auxiliary variables. This approach admits joint causal discovery with multiple treatments and multiple outcomes, which is particularly suitable for gene expressions studies, Mendelian randomization and EHR studies. The specificity score is related to Hill’s specificity criterion, but I will discuss the differences. Simulations are used for illustration and an application to a mouse obesity dataset detects potential active effects of genes on clinical traits that are relevant to metabolic syndrome.

Sensitivity Analysis

Sensitivity Analysis for Quantiles of Hidden Biases in Matched Observational Studies
Dongxiao Wu* Xinran Li,

In matched observational studies, the inferred causal conclusions pretending that matching has taken into account all confounding can be sensitive to unmeasured confounding. In such cases, a sensitivity analysis is often conducted. In general, a sensitivity analysis tries to infer the minimum amount of hidden biases needed in order to explain away the observed association between treatment and outcome, assuming that the treatment has no effect. If the needed bias is large, then the treatment is likely to have significant effects. The Rosenbaum sensitivity analysis is a modern approach for conducting sensitivity analysis for matched observational studies. It investigates what magnitude the maximum of the hidden biases from all matched sets needs to be in order to explain away the observed association, assuming that the treatment has no effect. However, such a sensitivity analysis can be overly conservative and pessimistic, especially when the investigators believe that some matched sets may have exceptionally large hidden biases. In this paper, we generalize Rosenbaum’s framework to conduct sensitivity analysis on quantiles of hidden biases from all matched sets, which are more robust than the maximum. Moreover, we demonstrate that the proposed sensitivity analysis on all quantiles of hidden biases is simultaneously valid and is thus a free lunch added to the conventional sensitivity analysis. An R package implementing the proposed method is also available online.

Sensitivity Analysis

Robust Causal Signatures from Joint Sensitivity Analyses of Multiple Estimands Nathan Cheng* Nathan Cheng, José Zubizarreta,

Design choices can improve the robustness of causal conclusions to unmeasured confounding. For instance, by using a negative control or by leveraging multiple estimands that are likely to be most sensitive to different kinds of confounding variation, sensitivity analyses can go beyond simple worst-case bounds and lead to much stronger causal conclusions than if these questions were treated separately. Although these ideas have been explored in the literature, the question of how one might accomplish this in a broader context—for instance, using more sophisticated estimators, or with the goal of producing multiplicity-adjusted sensitivity intervals—is relatively underdeveloped. In this work, we present an estimation framework, a flexible class of sensitivity models, and methods for producing valid sensitivity-adjusted hypothesis tests and confidence intervals that also account for the multiplicity of the estimands. We demonstrate how one can control sensitivity-adjusted versions of classical multiple testing error rates: the family-wise error, the false discovery/coverage proportion, and the false discovery/coverage rate. We illustrate gains in three particular settings: when a negative control outcome is available, when we have a multivariate treatment with a monotonic dose-response relationship, and when we have multiple not-too-perfectly-correlated outcomes.

Sensitivity Analysis

Universal Randomization Inference and Sensitivity Analysis for Matched Observational Studies with Continuous Treatments Jeffrey Zhang* Jeffrey Zhang, Siyu Heng, Dylan Small,

Matching is one of the most commonly used causal inference frameworks in observational studies. Through matching on measured confounders, valid randomization inferences can be conducted assuming no unmeasured confounding, and sensitivity analysis can be further performed to assess the sensitivity of randomization inference to potential unmeasured confounding. However, there is still a lack of valid randomization inference and sensitivity analysis approaches in many common matched studies. Specifically, in matched studies with continuous treatments, with the exception of special cases such as pair matching, there is no existing randomization inference or sensitivity analysis approach for studying analogs of the sample average treatment effect (Neyman’s weak null), and no existing valid sensitivity analysis approach for testing the null effect (Fisher’s sharp null) when the outcome is non-binary. To fill these gaps in matched studies with continuous treatments, we propose 1) a novel randomization inference and sensitivity analysis approach for studying weak null estimands, and 2) a valid sensitivity analysis approach for testing the null effect. Both of these two approaches work for general matching designs and general outcome variables. We illustrate our approaches via extensive simulations and a real-data application.

Synthetic Control Method

SMaC: Spatial Matrix Completion method Giulio Grossi* Giulio Grossi, Alessandra Mattei, Georgia Papadogeorgou,

Synthetic control methods are commonly used in panel data settings to evaluate the effect of an intervention. In many of these cases, the treated and control units correspond to spatial areas such as regions or neighborhoods. We work in a setting where a treatment is applied at a given location and its effect can emanate across space.
Synthetic control methods can be used to evaluate the effect that the treatment had in an area of a certain size around the treated location, but it is often unclear how this area should be defined, how far the treatment's effect propagates, or how the effect varies as a function of distance from the intervention point. Researchers might apply synthetic control methods separately for areas of different sizes around the treated location, but this approach ignores the spatial structure of the data and can lead to efficiency loss in spatial settings.
To address this, we develop a Bayesian spatial matrix completion framework that allows us to predict the missing potential outcomes in areas of different size around the intervention point while accounting for the spatial structure of the problem.
Similarly to synthetic control methodology, we impute the missing time series in the absence of treatment for areas around the intervention point using a weighted combination of control time series. In spatial problems, we expect areas of similar distance from the intervention point to be similar. Therefore, we impose that the imputation models vary smooth

Synthetic Control Method

A Synthetic Control Design for Moderate Covariate and Sample Size Lingke Jiang* Lingke Jiang, Yanran Li,

The synthetic control method (SCM) is an increasingly popular tool for analysis of policy efficacy and exposure effects in context of environmental health science. Here, it is applied to estimate the long-term effect of tropical cyclones on community social vulnerability in the U.S., which utilizes panel data containing over 20 years of time series. In this setting, it may not be sufficient past trajectories in the outcome of interest (as in the original SCM) but also demographic and meteorological histories between the control and treatment groups. However, while covariate balancing conditions may be ideal in settings of large number of covariates and sample size, such was not the case in our analysis. Based on the original SCM design, we extended the convex optimization algorithm in the original SCM such that it accommodates for a moderate number of covariates and sample size.

Weighting

Off-policy evaluation using debiased calibration weighting Yuyang Li* Jae-kwang Kim, Yumou Qiu, Yonghyun Kwon,

Calibration weighting is an important tool for improving efficiency of the design-based estimators in survey sampling. We propose a unified framework for debiased calibration in reducing selection bias and improving the efficiency of the estimator in the context of causal inference. Our approach is based on the generalized entropy as the objective function for optimization. The constraint incorporating the design weights is used to correct the selection bias while the benchmarking constraints are used to reduce the variance. The resulting calibration estimator is asymptotically equivalent to the design-optimal regression estimator of Deville and Sarndal (1992). We also propose the cross entropy as the optimal entropy function for calibration. To test our theory, we perform a limited simulation study.

Weighting

Semiparametric inference for sample treatment effects on the treated Andrew Yiu* Andrew Yiu,

Causal effects are most often defined in terms of a target superpopulation from which the observed data were drawn. This includes classic estimands such as the (population) average treatment effect on the treated (ATT). A possible alternative is to replace certain components of the population estimand with their sample analogues–for instance, we could define our causal effect by averaging over the observed covariates in the sample rather than over the unknown superpopulation covariate distribution. This could be implemented after the sample characteristics have been adjudged to be adequately representative of the target population. An advantage of these sample estimands is that we can estimate them more precisely than their population counterparts (i.e. smaller asymptotic variances, tighter confidence sets). Despite this, the theory for treatment effects on the treated is underdeveloped. We fill this gap by categorizing the family of sample estimands and establishing asymptotic properties with respect to a semiparametric efficient estimator. Our results include some surprising findings: the asymptotic variance of the popular “sample treatment effect on the treated” (SATT) is point-identified and can hypothetically exceed that of its population ATT counterpart; we also introduce a new estimand that can always be estimated at least as efficiently as both the ATT and the SATT.