Causal Inference Challenges and New Directions for Epidemiologic Research on the Health Effects of Social Policies

Ellicott C. Matthay[1] · M. Maria Glymour[2]

Abstract

Purpose of Review Epidemiologic research on the health effects of social policies is growing rapidly because of the potentially large impact of these policies on population health and health equity. We describe key methodological challenges faced in this nascent field and promising tools to enhance the validity of future studies.

Recent Findings In epidemiologic studies of social policies, causal identification is most commonly pursued through confounder-control but use of instrument-based approaches is increasing. Researchers face challenges measuring relevant policy exposures; addressing confounding and positivity violations arising from co-occurring policies and time-varying confounders; deriving precise effect estimates; and quantifying and accounting for interference. Promising tools to address these challenges can enhance both internal validity (randomization, front door criterion for causal identification, new estimators that address interference and practical positivity violations) and external validity (data-driven methods for evaluating heterogeneous treatment effects; methods for transporting and generalizing effect estimates to new populations).

Summary Common threats to validity in epidemiologic research play out in distinctive ways in research on the health effects of social policies. This is an active area of methodologic development, with ongoing advances to support causal inferences and produce policy-relevant findings. Researchers must navigate the tension between research questions of greatest interest and research questions that can be answered most accurately and precisely with the data at hand. Additional work is needed to facilitate integration of modern epidemiologic methods with econometric tools for policy evaluation and to increase the size and measurement quality of datasets.

Keywords Causal inference · Social policy · Quasi-experiment · Legal epidemiology · External validity · Econometrics

Introduction

Policies influencing social determinants of health are promising tools to improve population health and reduce health inequities. Legal rules established at any government jurisdiction (e.g., federal, state, city) can influence the distribution of social or behavioral determinants of health. Social policies include laws affecting education, income, immigration, labor, human rights, and employment, as well as

those that regulate products such as alcohol, tobacco, and firearms [1•, 2•]. Such social policies may have broader impacts than policies regulating medical care or health care financing [3–5]. Although social policies have the potential to enhance health equity, they can also sustain or propagate oppression and inequity. Rigorous evaluation of these policies is thus increasingly recognized as an important domain for epidemiologic research.

Research on the health effects of social policies draws on both epidemiology and social science. Substantive knowledge about social policies and methodologies for policy evaluation are often better represented in economics and sociology than epidemiology. Research on drivers of population health is the essence of epidemiology. This convergence brings new methodological challenges, and growing interdisciplinary methodological work is helping to bridge the language, methods, and evidence across these disciplines [6–8, 9•, 10–12].

✉ Ellicott C. Matthay
ellicott.matthay@ucsf.edu

1 Center for Health and Community, School of Medicine, University of California San Francisco, San Francisco, CA, USA

2 Department of Epidemiology and Biostatistics, School of Medicine, University of California San Francisco, San Francisco, CA, USA

Like all research aiming to draw causal inferences, studies on the health effects of social policies require strong assumptions and must address potential violations of conditional exchangeability, positivity, and consistency, among others (Box 1) [9•, 13]. However, policy studies face unique challenges to these assumptions. Randomization is frequently not feasible or ethical for policy evaluations, and researchers often seek to evaluate policies that have already been implemented in a non-randomized fashion. The small number of jurisdictions included in most policy studies (e.g., there are only 50 US states) often causes positivity violations. Interference is expected because people are influenced by the policies in jurisdictions neighboring their own, and because policies change social norms. Therefore, policy research routinely requires innovating on traditional epidemiologic tools to accommodate violations of the conventional assumptions, for example by modifying the target parameter and corresponding research question [14]. Yet not all studies on the health effects of social policies are designed and executed with careful attention to such methodological issues [1•]. Without appropriate consideration of the study design, statistical analysis, and interpretation, policy evaluations can be useless, or even harmful.

We begin this review by summarizing common approaches to studying the health effects of social policies. The subsequent sections elaborate how the assumptions for causal inference play out and are challenged for epidemiologic research on the health effects of social policies. Interwoven are promising methodologic frontiers in social policy research and tools for enhancing study validity.

Box 1. Definitions of key causal concepts

---

1. **Potential outcome**: The outcome that an individual (or other unit of analysis, such as family or neighborhood) would experience if his/her treatment (or exposure) takes any particular value. Each individual is conceptualized as having a potential outcome for each possible treatment value. Potential outcomes are sometimes referred to as counterfactual outcomes

2. **Exchangeability:** The assumption of no confounding, i.e., the assumption that which treatment an individual receives is unrelated to her potential outcomes if given any particular treatment. This assumption is violated for example if people who are likely to have good outcomes regardless of treatment are more likely to actually be treated. In the context of instrumental variables analysis, exchangeability is the assumption that the instrument does not have shared causes with the outcome

3. **Conditional exchangeability**: The assumption that exchangeability is fulfilled after controlling for a set of measured covariates. When this assumption is met, we say that the set of covariates—known as a **sufficient set**—fulfills the **backdoor criterion** with respect to the treatment and outcome

4. **Positivity**: All subgroups of individuals defined by covariate stratum (e.g., every combination of possible covariate values) must have a nonzero chance of experiencing every possible exposure level. Put another way, within every covariate subgroup, all exposure values of interest must be possible

5. **Consistency**: The assumption that an individual's potential outcome setting treatment to a particular value is that person's actual outcome if s/he actually has that particular value of treatment. This could be violated if the outcome might depend on how treatment was delivered or some other variation in the meaning or content of the treatment. Some researchers consider consistency a truism rather than an assumption

---

The informal definitions presented in this box are quoted from Matthay et al. 2019 [9•]

Common Approaches

The analytic approach is driven by both substantive interests and methodologic constraints. We distinguish between research evaluating the health effects of social policies by controlling confounders (Fig. 1a) and research that uses policies as instruments to quantify the effect of a social resource regulated by the policy (Fig. 1b) [9•]. For example, we might evaluate the effect of the earned income tax credit (EITC) policy on smoking prevalence [15], or we might instead use state-to-state variation in EITC benefit generosity to estimate how the extra income delivered by EITC affects health [16]. Both types of study are relevant to understanding policy effects. Evaluating the effects of a policy is useful for example to anticipate the effect of adopting a similar policy in the future. Using the policy as an instrument to evaluate the effects of the social resource regulated by the policy (i.e., defining that resource as the endogenous variable in an instrumental variable analysis) can strengthen causal evidence on social determinants of health and generate evidence to predict the impact of alternative interventions on those social resources. For example, EITC is one of the many possible policies to increase income for low-income families; one could also increase income by altering the minimum wage or enhancing Temporary Assistance to Needy Families benefits. Establishing the health benefits of extra income will help anticipate the consequences of these other income-support policies.

Figure 1 summarizes common confounder-control and instrument-based methods. All confounder-control methods are all premised on the assumption that the factors that determine policy exposure and also influence heath (or proxies for these factors) are fully measured and accounted for in the statistical analysis. For example, average education, income, or attitudes of community residents all affect health and may influence the likelihood a policy is adopted. Researchers often attempt to measure and control for each of these variables. In contrast, instrument-based methods require the assumption that the instrument (e.g., the policy adoption indicator) is unrelated to the health outcome that people in the sample would have experienced under alternative values of the policy exposure. Although instrument-based methods in health research more commonly evaluate the effect of a resource delivered by the policy, they may also evaluate a policy as the exposure.
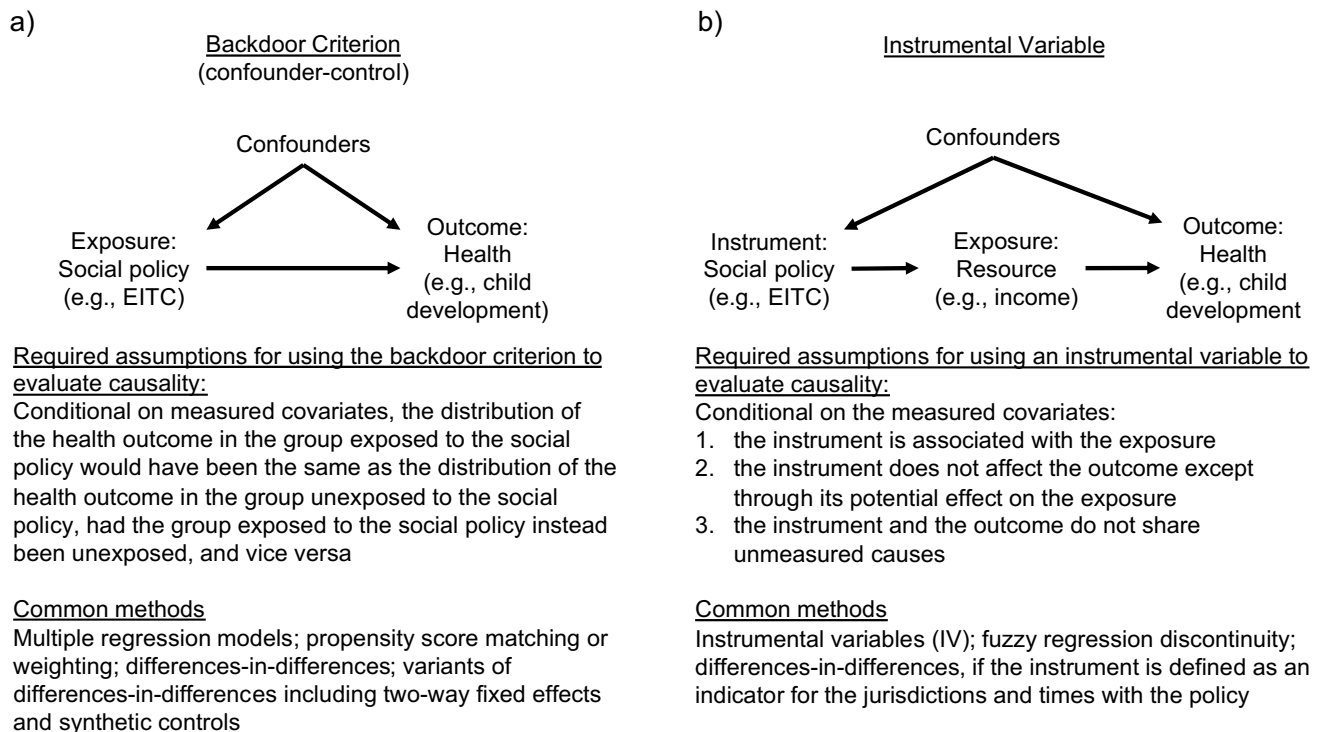
a)

Backdoor Criterion
(confounder-control)

Confounders

Exposure:
Social policy
(e.g., EITC)

Outcome:
Health
(e.g., child
development)

Required assumptions for using the backdoor criterion to evaluate causality:
Conditional on measured covariates, the distribution of the health outcome in the group exposed to the social policy would have been the same as the distribution of the health outcome in the group unexposed to the social policy, had the group exposed to the social policy instead been unexposed, and vice versa

Common methods
Multiple regression models; propensity score matching or weighting; differences-in-differences; variants of differences-in-differences including two-way fixed effects and synthetic controls

b)

Instrumental Variable

Confounders

Instrument:
Social policy
(e.g., EITC)

Exposure:
Resource
(e.g., income)

Outcome:
Health
(e.g., child
development

Required assumptions for using an instrumental variable to evaluate causality:
Conditional on the measured covariates:
1. the instrument is associated with the exposure
2. the instrument does not affect the outcome except through its potential effect on the exposure
3. the instrument and the outcome do not share unmeasured causes

Common methods
Instrumental variables (IV); fuzzy regression discontinuity; differences-in-differences, if the instrument is defined as an indicator for the jurisdictions and times with the policy

Fig. 1 Directed acyclic graphs and required assumptions for evaluating causation via the backdoor criterion versus an instrumental variable approach Legend: Social policy research may evaluate the effects of the policy per se. These research questions are typically approached by controlling for confounders that influence policy adoption and health or by controlling for fixed effects for places that adopt the policy at different times (Fig. 1a). Alternatively, policies may be used as instruments to estimate the effect of the resource regulated by the policy, such as income, on health (Fig. 1b). Either approach requires strong assumptions to support causal inferences. For instrumental variable approaches, the three assumptions listed are required for evaluating causality, but causal identification also usually relies on a fourth assumption of monotonicity—that is, that the instrument does not affect the likelihood of exposure in opposite directions for different people in the sample. This figure provides intuitive informal definitions of the required assumptions for evaluating causality. There are many different ways to state the formal required assumptions, and we refer the reader to Hernán and Robins [13], Pearl [17], and Glymour and Swanson [18] for different variations. In particular, one alternative way to state the assumptions is using potential outcomes language. For the backdoor criterion, the required assumption to evaluate causality can be phrased as: Conditional on measured covariates, the potential health outcome is unrelated to the actual level of the social policy exposure. Likewise, the required assumption for using an instrumental variable to evaluate causality is: Conditional on measured covariates, the instrument is unrelated to the potential health outcomes that people in the sample would experience under alternative values of the exposure.

Instrument-based methods are commonly called quasi-experiments because the variation in the exposure that is induced by the instrument is thought to be like-random or arbitrary. Instrument-based methods are commonly perceived as having greater internal validity, because they circumvent the need to correctly identify, measure, and control for a set of covariates sufficient to control all confounding [9•]. However, data from quasi-experimental settings can be statistically analyzed using either confounder-control or instrument-based methods, depending on the substantive question. The internal validity of the design derives from the quasi-experimental context, not from the analytic choice.

Since the publication of foundational work on the causal interpretation of instrument-based methods [19], IV has emerged as a hallmark of applied econometrics. In epidemiologic research involving social policies, confounder-control methods are more common, but uptake of instrument-based methods is increasing (Fig. 2) [20]. For both instrument-based and confounder-control methods, meeting the required assumptions for conditional exchangeability is challenging. Not every policy change is a valid instrument for a measured exposure variable. Multiple policies may change simultaneously. Places that adopt a policy may differ in important and unmeasured ways from places that do not adopt the policy. Thus, policy studies are most compelling when the source of quasi-experimental variation is plausibly like-random or arbitrary (Table 1).

What effects are measured and among whom? Measurement, consistency, and spillover

In research on the health effects of social policies, *identifying and measuring relevant policy exposures* is a major
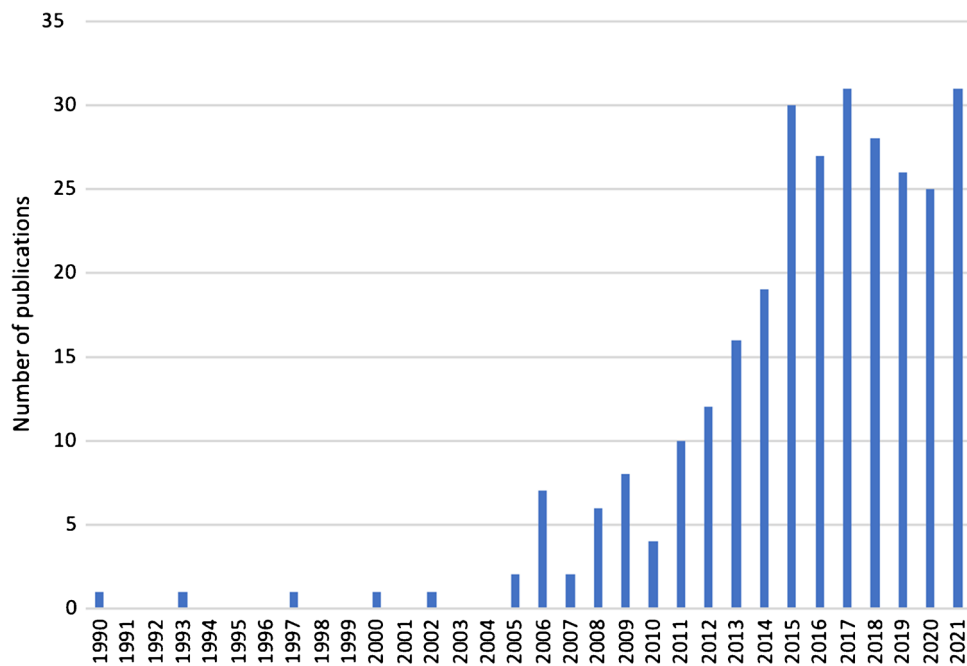
Fig. 2 Number of publications mentioning quasi-experimental methods in the title or abstract in selected Epidemiology journals, 1990–2021 Legend: Data generated from PubMed search "Results by Year" feature using search term: (("instrumental variable"[Title/Abstract] OR "quasi-experiment"[Title/Abstract] OR "natural experiment"[Title/Abstract] OR "difference-in-difference"[Title/Abstract] OR "differences-in-differences"[Title/Abstract] OR "synthetic control"[Title/Abstract] OR "panel fixed effects"[Title/ Abstract] OR "two-way fixed effects"[Title/Abstract] OR "regression discontinuity"[Title/Abstract])) AND ("American Journal of Epidemiology"[Journal] OR "Epidemiology"[Journal] OR "International Journal of Epidemiology"[Journal] OR "Journal of Epidemiology and Community Health"[Journal] OR "Annals of Epidemiology"[Journal] OR "European Journal of Epidemiology"[Journal])

challenge. Policies are made up of diverse provisions and come in many forms, including statutes, ordinances, regulations, case law, ballot initiatives, and referenda [61•]. Specifying the types of policies assessed in a given study is critical to ensure replicability and clarify the scope. Complex overlays of city, county, state, federal, and international jurisdictions mean that establishing the legal framework—i.e., what can be regulated and by whom—is non-trivial [61•]. Understanding these complexities usually requires collaboration with legal experts [61•, 62]. Legal text must be translated into numeric codes for quantitative research. Determining which aspects of policies matter for health, and thus which ones should be measured is a challenge [61•]. Furthermore, written policies may not correspond to what is actually being implemented or enforced [63, 64]. Legal Epidemiology resources are available to guide this process and maximize rigor and replicability. Efforts are growing to support theory-based policy taxonomies [65•, 66], legal databases that are systematically coded and regularly updated (e.g., Law Atlas Policy Surveillance Program, Alcohol Policy Information System), and quantification of policy enforcement [63, 64]. Qualitative work is essential to fully understand how policies operate and affect individuals.

Most policy evaluations use binary indicators of policy enactment, but more nuanced and multi-valued measures may better address threats to positivity and conditional exchangeability [67•]. Alternatives to binary measures include the proportion of the population eligible for or receiving a policy benefit [68]; the generosity of benefits [33]; the size of a tax, subsidy, or penalty [33, 69, 70]; the magnitude of uptake of a newly permitted activity (e.g., the number of new cannabis dispensaries) [71, 72]; or the degree of enforcement [73]. If it is not possible to identify the effect of a single policy, the joint effects of policies adopted as a bundle [74, 75] or the effect of the policy environment as measured by overall policy stringency or comprehensiveness [34, 35, 76] may be identifiable.

Analyses involving distinct exposure measures estimate different parameters and therefore answer different research questions. Researchers should consider what information will be most useful for developing new policies. Policymaking is notoriously messy and unpredictable, so the precise policy implemented in the past may not be politically achievable in the future. Research to understand the effects of specific policy features, and evaluations that can offer broader theoretical insights, are often more informative than research to understand a single, narrowly defined policy instantiation.

Table 1 Potential sources of quasi-experimental variation for studies of social policies

Source of arbitrary or like-random variation	Examples
Lotteries	Wartime draft lottery [21], housing vouchers [22], school vouchers [23], randomized refugee dispersal [24]
Discontinuities based on dates of policy adoption or implementation	Minimum unit pricing on alcohol purchases [25], lowering the blood alcohol content limit for drivers [26], sugar-sweetened beverage taxes [27]
Region and time variation in policy adoption or implementation	Electoral cycles [28], immigration laws [29], motor vehicle safety laws [30], paid family leave laws [31]
Timing of delivery of benefits of a policy	Tax credits and short-term health outcomes in months on tax disbursements (February, March, April) [15], monthly Supplemental Nutrition Assistance Program disbursements and health outcomes in first half versus second half of the month [32]
Policy intensity, restrictiveness, or generosity	Earned income tax credit benefit generosity [16], unemployment benefit generosity [33], firearm law restrictiveness [34], alcohol law restrictiveness [35], minimum wage level [36], beer excise tax [37]
Physical proximity to a jurisdiction with a policy or a resource delivered by a policy	Proximity to educational institutions [38], proximity to borders of states with cannabis legalization [39, 40], proximity to cardiac care centers [41]
Cutoffs or discontinuities in eligibility for resources delivered by a policy based on age, income level, or similar factors	Handgun purchase age minimums [42], education policy changes (affected versus unaffected cohorts) [43, 44], Supplemental Nutrition Assistance Program (participants versus non-participants) [32, 45], cash transfers (recipient versus non-recipient) [46, 47], Deferred Action for Childhood Arrivals (eligible versus ineligible) [48]
Biological chance	Month or quarter of birth [49, 50], family size or sibling composition [51, 52], genetic variants (Mendelian Randomization) [53, 54]
People who move residence across jurisdictions	Longitudinal studies of the impact of the built environment among movers and non-movers [55, 56]
Assignment to service providers or officials	Clinicians who have different preferences for treatment modalities [57, 58], judges who have different propensities for leniency [59, 60]

Conflicting evidence on the effects of a policy may result from differing exposure measures. These can be conceptualized as due to *violations of the consistency assumption* [77]. Varying approaches to implementing similar policies in different places can result in different policy impacts. For example, studies of cannabis legalization often use a binary indicator of legalization, but health impact depends on whether the legalization policy enables retail sales [78–80]. For IV studies of the effects of a resource delivered by a policy, it is important to consider whether the consistency assumption holds with respect to the resource variable—for example, would the health benefit of extra income from EITC be similar to the health benefit from extra income due to minimum wage increases.

To avoid consistency violations, all of the relevant provisions and components of implementation must be correctly identified and measured—a particularly challenging task when it is not yet known what matters. Consistency is also problematic for policy measures that are composites, sums, or scores. Policy analyses often treat a one-unit increase in the score as having the same effect regardless which policy was added or the baseline level of the score [34, 81], yet not all policies are equally effective for all outcomes. When consistency violations occur, the evidence delivered by a policy

study will not clearly indicate what needs to be done (or not done) to replicate the study's findings, and policy decisions motivated by this evidence may not yield the desired results.

Social policies often have different effects for different population subgroups—i.e., *heterogeneous treatment effects* (HTEs) [82•]. For example, tobacco clean air policies especially benefit people with low levels of education [83] while paid family leave policies may improve breastfeeding outcomes primarily for high-income mothers [31]. HTEs are important to quantify for several reasons. Policies that disproportionately benefit the vulnerable can reduce health inequities, whereas the reverse can exacerbate inequities. Quantifying HTEs also supports research to anticipate whether the effect of an intervention will differ if implemented in a new population with a different composition. For example, recent extensions of methods for generalizing or transporting effect estimates can account for differences in the socioeconomic status, demographics, clinical profiles, or other characteristics in new settings [84•, 85–88]. On its own, understanding policy effects for specific subgroups (e.g., those receiving the intervention) may be a goal of the research [89, 90]. Alternatively, HTE research can indicate who is most likely to have the largest benefit from a given resource, a priority for decision-makers with limited resources.

We distinguish methods for evaluating HTEs based on a priori specified characteristics from data-driven methods which more agnostically search for groups with heterogeneous responses to the policy [82•]. Heterogeneity across pre-specified characteristics (e.g., race/ethnicity, age) is usually quantified via stratification or interaction terms in statistical models, and the chosen dimensions can be guided by theory or evidence [82•]. Data-driven methods can be used to evaluate whether there is *any* heterogeneity across any characteristics (e.g., across all possible combinations of covariates), to partition the participants into subgroups that have different policy responses, to identify the most-affected subgroup, or to identify optimal policy combinations [67•, 91–95]. Data-driven HTE methods are one of the many emerging applications of machine learning in policy research, but they remain rare in applied health-related policy studies [67•, 82•].

Existing guidance on how HTEs should be evaluated, reported, and interpreted has been limited to randomized trials and clinical applications [96–100]. Policy studies involving HTEs require additional guidelines. Testing for policy effects in multiple population subgroups increases the risk of spurious findings, especially when considering subgroups defined by multiple covariates simultaneously. How this risk should be weighed against the knowledge gained and which methods are most appropriate for limiting spurious findings in policy contexts are not established. Precision concerns are exacerbated when examining HTEs. Given the limited resources, HTE evaluation should be prioritized for policies, settings, and population subgroups for which HTEs are likely to be substantial enough to alter recommendations for policy or practice. Yet few detailed empirical studies or theoretical frameworks are available to guide prioritization. Tools to robustly evaluate HTEs in small sample sizes are also needed.

In policy studies, *unexpected, inconsistent, and null results* are common. The Moving to Opportunity housing experiment benefitted girls and harmed boys, at least for some outcomes [22]. "Ban the box" policies, designed to reduce racial disparities in employment by preventing employers from conducting criminal background checks at certain stages of hiring, paradoxically exacerbated inequalities for low-skilled workers [101]. Identifying the causes of discrepant results can be challenging because the data sources, measures, statistical methods, and settings of policy studies are so diverse [1•]. Implementation science and qualitative research can enhance social policy research by unveiling what went wrong or what needs to be done differently (e.g., solutions to differential uptake). Policy research is often disconnected from *community engagement* efforts, but such research will be stronger and more pertinent if researchers build in strategies to involve and solicit feedback from communities affected by the policies and health outcomes under study.

Spillover or interference violates the independence assumption of standard statistical approaches in social policy research and can lead to anticonservative standard errors and spurious associations. Policies may change health not only for the people to whom the policies directly apply but also for family members, neighbors, or other social contacts—for example, because the health outcome is contagious, responsive to social norms (e.g., smoking), or social in nature (e.g., violent injury). Neighboring jurisdictions with differing laws may motivate individuals to cross borders to avoid or pursue specific policies. For example, cross-border influences have been investigated for policies regulating firearms [102, 103], tobacco [104], and sugar-sweetened beverages [105]. Policies operating at lower levels of aggregation (e.g. cities versus countries) may be more susceptible to spillover, because individuals with differing policy exposures are more likely interact or cross borders. Spillover and interference can magnify the effects of social policies (e.g., via contagion) or attenuate effects (e.g., crossing borders to avoid the policy). Either phenomenon threatens accurate estimation of causal effects.

Potential solutions include spatiotemporal modeling, complex systems modeling, social network analyses, and estimating novel parameters. Researchers have used Bayesian spatiotemporal analyses with conditional autoregressive random effects to account for spatial autocorrelation in studies of local alcohol and cannabis policies [106, 107] and agent-based models have been used to evaluate the impact of past and potential alcohol and opioid policies (see Appendix for more on complex systems modeling) [108, 109]. Estimating parameters that explicitly incorporate dependencies between jurisdictions or groups in their definition [110] or that allow for policy exposures to be assigned stochastically rather than deterministically [111] can also help address challenges to spillover and interference. Additionally, novel econometric work is advancing methods for assessing causal effects in social networks (e.g., coworkers, classrooms, or neighbors) [112, 113].

Generalizing study results to new populations is almost always a goal of epidemiologic research. Advances in theory and statistics increasingly support methods to generalize or transport effect estimates [84•, 85–88]. Given that social policy effects likely differ across population subgroups, a given policy may have very different impacts in one population versus another. Methods for transport and generalization allow researchers to predict the impact of a policy in a population that is different in composition from the one initially studied. Despite the relevance of these methods to policy studies, transportability estimators have almost exclusively been applied to research based in randomized trials. All studies involve tradeoffs between internal and external

validity, and no single study can achieve the optimal degree of internal and external validity [9•, 114•]. Even with imperfect internal validity, observational and quasi-experimental studies aiming to generalize estimates of policy effects to new populations can be valuable.

Statistical considerations and data needs

Policies often have small effects on individuals [115•]. Although small effects can matter immensely when applied to an entire population [116], many social policy studies lack sufficient sample sizes to precisely estimate these effects [117•]. *Underpowered studies* risk concluding that a health-promoting or harmful policy has no effect. For this reason, large-scale administrative health data such as vital statistics or Medicare billing records are commonly used for policy studies. Electronic health records (EHR) and other clinical information systems are increasingly viable possibilities for policy analysis. The Affordable Care Act accelerated uptake of EHRs and incentivized adoption of a common data framework, although the lack of population representativeness in such datasets has not been fully addressed.

Administrative data tend to have less detailed, *lower quality measurements*, and rarely contain information on both social policy exposures *and* health outcomes. Therefore, health-related social policy evaluations often require linkages across multiple data sources or sacrificing measurement quality. Residential address can be used to link individual-level health outcomes to relevant policy jurisdictions. Residential history information even permits linkage across life course periods. Few datasets offer comprehensive residential histories, but incorporating or linking residential histories into existing datasets would enhance policy research.

Given the limitations of administrative data, potential strategies to improve precision and measurement quality include taking detailed measurements on random subsamples of large datasets [118]; incorporating measures of social program participation into existing large-scale health data collection [119]; incorporating improved health measures into large-scale generalized datasets (e.g., American Community Survey); and supporting big data initiatives that harmonize multiple individual-level, geographically detailed administrative datasets [120]. Additionally, Bayesian statistical methods are rare in policy studies [1•], but these methods can enhance precision by drawing on prior knowledge about policy effects and bounding the range of plausible effect estimates.

If the sample size is not fixed, power calculations can help ensure that a study is designed to achieve sufficient precision. In practice, power calculations are uncommon for policy studies with existing data, and in retrospect may have very low power [117•, 121, 122]. Over-reliance on null-hypothesis significance testing is a pernicious problem in policy analyses. Interpreting the confidence interval for the estimated effect is essential to avoid concluding that an underpowered study demonstrates that the policy has no important effects. If the confidence interval for the policy effect crosses the null but includes values that would be of substantial benefit or harm when applied at a population level, the study is underpowered.

Internal validity, conditional exchangeability, and positivity

Confounding bias arises from systematic differences between jurisdictions with different levels of policy exposure (e.g., states that did and did not adopt a given policy). The social, political, and economic forces that shape policies also affect many health outcomes. Confounding can be severe and intractable—for example, when the confounders and the policy are too closely aligned to be disentangled. Strong confounders of policies include other policies and jurisdiction-level political orientations, especially for polarizing issues such as firearms, abortion, or immigration [2•]. For example, the restrictiveness of state firearm policies is strongly determined by state political orientations, and states with stricter policies tend to have many policy restrictions, making it difficult or impossible to disentangle the effect on any one policy from the others [2•, 123]. Quantitative bias analysis, negative control exposures and outcomes, and other robustness checks are valuable yet underutilized tools for assessing the likely direction and magnitude of confounding bias in social policies studies (see Appendix for detail). For these tools, a high priority is to standardize their inclusion in policy studies. Potential checks should be articulated even when they cannot be fielded in a particular dataset.

Most policy studies involve longitudinal data structures and the possibility of *time-varying confounding*. Time-varying confounding can occur when prior levels of a policy exposure affect downstream confounders which in turn affect subsequent policies and health outcomes. For example, US state prescription drug monitoring programs (PDMP) adopted in response to the overdose crisis may have affected illicit opioid market dynamics by decreasing access to prescription opioids and increasing demand for illicit opioids [124, 125]. This pattern may prompt further changes to state opioid policies to reduce use of illicit opioids in an attempt to reduce overdose deaths. For studies evaluating the effects of opioid policies over time, illicit opioid market dynamics partially mediate the PDMP-overdose relationship, but confound the relationship between late-stage opioid policies and overdose. Bias therefore results from both typical regression adjustment for illicit opioid market dynamics or failure to adjust for illicit opioid market dynamics. Other examples of potential time-varying confounders include participation in social programs (e.g., SNAP, WIC, EITC, etc.), receipt of

physical or mental health services, smoking status, alcohol use, exposure to air pollution, access to green space, residence in public housing, and diet quality.

Multiple methods have been developed to address time-varying confounding including inverse probability weighted estimation of marginal structural models, g-estimation of structural nested models, the longitudinal g-formula, and longitudinal targeted minimum loss-based estimation [126–128]. However, these methods are rarely used in applied studies, including to evaluate social policies [129]. Barriers to uptake of these methods include uncertainty about how to integrate them with IV or DID analyses of quasi-experiments and data requirements (large sample sizes, repeated observations on units over time).

Multiple related policies are often adopted or implemented in the same jurisdiction simultaneously or in quick succession, a problem known as *co-occurring policies* [1•, 2•]. Many study designs exploit variation in the timing and locations of policy changes across jurisdictions to isolate the causal effects of a policy. Co-occurring policies that all affect the outcome of interest pose a conundrum for such designs: left uncontrolled, co-occurring policies confound one another, but controlling for co-occurring policies can reduce effective sample size and lead to positivity violations. Positivity violations can lead to bias, imprecision, and undefined estimates [1•, 2•]. This challenge is pervasive across numerous social policy domains and often results in very imprecise effect estimates [2•]. Potential solutions to co-occurring policy problems include explicitly assessing threats to positivity, restricting to population subgroups among whom the policies can be disentangled, using more nuanced measures of policy exposure, defining clusters of policies as the exposure(s) of interest, or using stringency or generosity scores to characterize the overall policy environment [1•, 14].

Randomization of policies is often considered unethical or impractical due to limited resources or political urgency, but experiments of public social interventions are often feasible. A recent systematic review identified 38 US social policies that were evaluated with randomized designs and examined health outcomes [117•]. Public and political support for social experiments may have waned in recent decades [130]. Epidemiologists have a role to play in advocating for the practicality, ethics, and benefits of randomization for science and public health. While legitimate ethical concerns exist, adoption and maintenance of potentially harmful policies and failure to randomize given the opportunity are also unethical.

New social policies can rarely be implemented instantaneously among all potential beneficiaries, and the need for gradual scale-up creates ethical opportunities to randomize at either the place-level (e.g., randomized stepped wedge designs [131]) or the individual-level (e.g., waitlist controls [132], lotteries [22]). The evaluation of California's Armed and Prohibited Persons System constitutes one of the largest-known cluster-randomized trials of a public policy [133•]. This initiative aimed to recover firearms from people who purchased them legally but later became prohibited from owning them. Collaboration between academic investigators and the California Department of Justice facilitated a randomized rollout of the program to the 1000 + communities who received intervention earlier or later.

Virtually all current efforts at inferring the causal effects of social policies on health can be characterized as either confounder-control or instrument-based methods [9•]. However, Pearl's transdisciplinary causal inference framework recognizes a third approach to causal identification: the *front door criterion*. This rarely used approach is premised on enumerating and measuring *all* the causal pathways by which an exposure affects an outcome (Fig. 3). If there are no unmeasured confounders of the exposure-mediator or mediator-outcome relationships, then an unbiased estimate of the causal effect of the exposure on the outcome can be quantified. Importantly, applications of the front-door criterion do not require the investigator to identify, measure, and appropriately adjust for confounders of the exposure-outcome relationship (i.e., backdoor adjustment). Thus, in social policy contexts where accurate and comprehensive confounder-control is challenging, and valid instruments are unavailable, but the mechanisms by which a policy affects health can reasonably be hypothesized and measured, the front door criterion shows great promise.

Glynn and Kashin demonstrated a compelling application of the front door criterion to estimate the effect of the often-studied Job Training Partnership Act (JTPA) program [134, 135] on subsequent earnings by leveraging program adherence as the mediator [136•]. Using data on participants and observational population-based controls, the authors evaluated varying specifications of front door and backdoor adjustment in comparison with results from the original JTPA randomized trial. While estimates from backdoor adjustment were sensitive to the choice of adjustment variables and often failed to replicate the trial results, the front door estimates consistently replicated trial results irrespective of the choice of adjustment variables. This study adds to a small but growing literature suggesting that front door approaches may be more robust to violations of the conditional exchangeability assumption than backdoor approaches [136•, 137, 138]. Similar applications can be conceptualized for social policies when the exposure is eligibility for resources and the mediator is program uptake or adherence.

Estimation methods that *move beyond traditional regression coefficients* also show promise in addressing threats to validity. For policy domains with political polarization, in which certain jurisdictions are unlikely to ever adopt a policy, defining the causal effect of interest not as an all-or-nothing contrast (e.g., all states adopt the policy versus no states adopt the policy) but rather as a temporal shift (e.g., what if all states adopting the policy delayed adoption by

Front Door Criterion
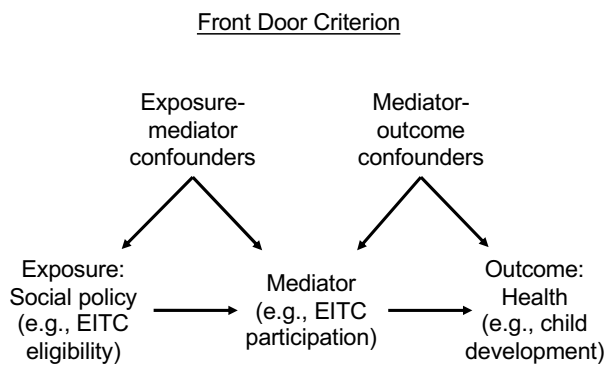
Required assumptions for using the front door criterion to evaluate causality:
1. No unmeasured confounders of the exposure-mediator relationship
2. No unmeasured confounders of the mediator-outcome relationship
3. The measured mediator(s) fully mediates the effect of the exposure on the outcome (i.e., there is no direct effect of the exposure on the outcome and no unmeasured mediators of the effect of the exposure on the outcome)

Fig. 3 Directed acyclic graph and required assumptions for evaluating causality via the front door criterion Legend: In the absence of a randomized trial, causality is typically evaluated via one of two approaches: fulfilling the backdoor criterion (i.e., controlling for confounders) or using an instrumental variable. A rarely used alternative is to evaluate causality by fulfilling the front door criterion. Using the front door criterion is appealing because it can be fulfilled even when there are unmeasured confounders of the exposure-outcome relationship. This figure provides intuitive informal definitions of the required assumptions for evaluating causality via the front door criterion. There are many different ways to state the formal required assumptions, and we refer the reader to Hernán and Robins [13] and Pearl [17] for different variations.

2 years) can result in a definition of positivity that is more achievable [139, 140•]. Recent advances in econometrics showed that the standard two-way fixed effects design—involving panel data on multiple jurisdictions over time and indicator variables for each jurisdiction, each time step, and exposure to the policy—can be substantially biased when: (1) the timing of policy implementation is staggered across jurisdictions; and (2) the effect of the policy within a jurisdiction changes over time [141•]. Both of these criteria are common for social policies and health. Multiple new estimators were developed to address this bias, but they have not yet been widely adopted in applied research [141•].

Conclusions

Rigorous research on the health effects of social policies can be highly relevant to public decision-making. Thus, epidemiologic research in this field is likely to grow. In the existing literature, study designs, data sources, policy domains, health outcomes, and populations of interest vary widely. Study quality also varies, and the strongest studies are those involving careful selection of the research question and precise attention to which populations represent the counterfactual outcomes, i.e., which populations are being used to approximate the outcomes the population exposed to the policy would have experienced if the policy had not been implemented. The estimand most plausibly identifiable with the data at hand is often not the estimand of greatest substantive interest. Navigating this tension—between the question of greatest interest and the question that can actually be answered—is a central focus of designing rigorous policy studies.

Numerous methodological tools are underutilized and are poised to enhance the rigor of research on the health effects of social policies as these methods become more accessible. Methodological development to integrate epidemiologic and econometric methods, and to increase the size and measurement quality of datasets will also improve study quality. Growing focus on evaluating heterogeneity in policy effects and anticipating the health effects of policies in new populations will expand the relevance of policy research and help to meet the growing demand for evidence to guide social policy decisions.

Appendix

Common confounder-control and instrument-based methods: Common confounder-control methods for policy research include multiple regression models [46, 76, 142, 143], matching or propensity score weighting [144, 145], many implementations of differences-in-differences (DID), and DID variants including two-way fixed effects and synthetic controls [26, 27, 29, 36, 146•, 147]. These methods are all premised on the assumption that the factors that determine policy exposure and also influence heath (or proxies for these factors) are fully measured and accounted for in the statistical analysis. For example, average education, income, or attitudes of community residents all affect health and may influence the likelihood a policy is adopted. Researchers often attempt to measure and control for each of these variables.

Instrument-based methods for policy research use instrumental variables (IV) analysis [45, 148, 149]. IV can also be implemented as fuzzy regression discontinuity [150–152], or DID, the latter if the instrument is defined as an indicator for the jurisdictions and times with the policy [153]. Instrument-based methods require the assumption that the instrument (e.g., the discontinuity indicator or the policy adoption indicator) is unrelated to the health outcomes that people in the sample would have experienced under alternative values of

the policy exposure. Although instrument-based methods in health research are commonly used to evaluate the effect of a social resource delivered by the policy, they may also be used to evaluate a policy as the exposure.

Power calculations: If the sample size is not already fixed, power calculations can help ensure a study is designed to achieve sufficient precision. However, as inputs to such calculations, evidence to guide the selection of anticipated effect sizes for social policies and health is sparse. Several considerations suggest large effect sizes are unlikely [115•]. First, most social policies act on health through intermediate levers such as housing or economic security. Second, a large fraction of the population may be ineligible or otherwise unaffected by the social policy; population-level impacts will be muted proportional to the fraction of the population for whom the policy is irrelevant. Some research suggests that power calculations are rarely done for policy studies with existing data and fixed sample sizes, and in retrospect may have very low power [117•, 121, 122].

Quantitative bias analysis, negative control exposures and outcomes, and other robustness checks are valuable yet underutilized tools for assessing the likely direction and magnitude of biases in social policies studies. Quantitative bias analyses (QBA) are a class of methods used to estimate the direction, magnitude, and uncertainty of systematic biases in estimated measures of association. For example, researchers have quantified the E-value, which indicates the minimum strength of associations that an unmeasured confounder would have to have with both the policy and the outcome to fully explain away the estimated policy-outcome association [103, 154]. In policy studies, QBA may be useful for example in estimating the potential magnitude of bias arising from a particular strong confounder or mis-measurement of relevant policy features. Guidance for conducting QBA is growing [155, 156, 157•], but debate remains over which approaches are most informative [158, 159].

Negative controls are useful for detecting biases including unmeasured confounding, recall bias, and analytic flaws [160]. For example, suicides completed with means other than firearms have been used as a negative control when evaluating the impacts of firearm policies on firearm suicides [161]. If firearm and nonfirearm suicides have similar determinants, but nonfirearm suicides are unaffected by firearm policies, then testing the association of firearm policies with nonfirearm suicides can be used to detect residual confounding due to these factors. Negative control variables can take many forms, and their utility depends on identifying a compelling negative control exposure or outcome. Other forms of robustness checks common in the econometric literature have been reviewed elsewhere [67•]. For research on the health effects of social policies, a high priority is to standardize the inclusion of QBA, negative controls, and other robustness checks. Potential checks should be articulated even when they cannot be fielded in a particular dataset.

Complex systems modeling, a suite of mathematical tools including compartmental, systems dynamic, and agent-based modeling, can help address the multi-level, dynamic, interrelated nature of real-world phenomena that make studies of social policies challenging [109, 162•]. These approaches can explicitly model feedback loops and reverse causation in which a policy is adopted in response to a problem, which can otherwise give the impression that a policy magnified the problem it was designed to address [124]. Systems modeling approaches can be used to answer causal questions about the impacts of social policies, but a main contribution of systems modeling for policy studies is predicting the potential impacts of hypothetical future policies in places where the policies have not yet been implemented [162•]. Lack of widespread standardized training in complex systems modeling is a primary barrier to greater uptake of these methods [163].

Another recent and exciting methodological advance for policy research is the development of *automated search algorithms* to support causal identification. For example, given a dataset, machine learning can be used to discover instrumental variables or natural experiments that can be leveraged for causal identification [164]. Causal search algorithms can also be used to determine which directed acyclic graphs are consistent with an observed dataset and thus which causal mechanisms are most plausible [165]. These emerging algorithms are promising both because of their potential to strengthen causal inferences about the health effects of social policies and to better understand the complex mechanisms via which social policies act to affect health. Pairing these algorithms with deep substantive knowledge is critical to ensure their appropriate use.

Declarations

Conflict of Interest The authors declare no competing interests.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as: ● Of importance

1.● Matthay EC, Gottlieb LM, Rehkopf D, Tan ML, Vlahov D, Glymour MM. What to do when everything happens at once: analytic approaches to estimate the health effects of co-occurring social policies. Epidemiologic Reviews [Internet]. 2021 [cited 2021 Aug 18]; Available from: https://doi.org/10.1093/epirev/mxab005. **A review of methods used for causal identification and to address co-occurring policies in contemporary studies of the health effects of social policies.**

2.● Matthay EC, Hagan E, Joshi S, Tan ML, Vlahov D, Adler N, et al. The revolution will be hard to evaluate: how co-occurring policy changes affect research on the health effects of social policies. Epidemiol Rev. 2021;mxab009. **A methodological review documenting the pervasiveness and consequences cooccurring policies across diverse social policy domains**.

3. Carey G, Crammond B. Systems change for the social determinants of health. BMC Public Health. 2015;15:662.

4. Galea S, Tracy M, Hoggatt KJ, DiMaggio C, Karpati A. Estimated deaths attributable to social factors in the United States. Am J Public Health. American Public Health Association; 2011;101:1456–65.

5. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. JAMA. 2004;291:1238–45.

6. Gunasekara FI, Carter K, Blakely T. Glossary for econometrics and epidemiology. J Epidemiol Community Health. 2008;62:858–61.

7. Duncan G j. When to promote, and when to avoid, a population perspective. Demography. 2008;45:763–84.

8. Kim Y, Steiner PM, Hall CE, Su D. Graphical Models for Quasi-Experimental Designs [Internet]. Society for Research on Educational Effectiveness; 2016 [cited 2018 Jul 9]. Available from: https://eric.ed.gov/?id=ED567736

9.● Matthay EC, Hagan E, Gottlieb LM, Tan ML, Vlahov D, Adler NE, et al. Alternative causal inference methods in population health research: evaluating tradeoffs and triangulating evidence. SSM - Population Health. 2020;10:100526. **An accessible introduction to the two main approaches to causal inference used in non-randomized studies of the health effects of social policies (confounder-control and instrument-based methods), with a glossary cross walking commonly used terms in econometrics and epidemiology**.

10. Matthay EC, Glymour MM. A graphical catalog of threats to validity. Epidemiology. 2020;31:376–84.

11. Shadish WR. Campbell and Rubin: a primer and comparison of their approaches to causal inference in field settings. Psychol Methods. 2010;15:3–17.

12. Krieger N. Epidemiology and social sciences: towards a critical reengagement in the 21st century. Epidemiol Rev. 2000;22:155–63.

13. Hernán MA, Robins JM. Causal inference: what if. Boca Raton: Chapman & Hall/CRC; 2020.

14. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res. 2012;21:31–54.

15. Rehkopf DH, Strully KW, Dow WH. The short-term impacts of Earned Income Tax Credit disbursement on health. Int J Epidemiol. 2014;43:1884–94.

16. Hamad R, Rehkopf DH. Poverty and child development: a longitudinal study of the impact of the Earned Income Tax Credit. Am J Epidemiol. 2016;183:775–84.

17. Causality PJ. New York. NY: Cambridge University Press; 2009.

18. Glymour MM, Swanson SA. Instrumental variables and quasi-experimental approaches. Modern Epidemiology. Fourth. Philadelphia, PA: Wolters Kluwer; 2021.

19. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc. 1996;91:444–55.

20. Bärnighausen T, Røttingen J-A, Rockers P, Shemilt I, Tugwell P. Quasi-experimental study designs series-paper 1: introduction: two historical lineages. J Clin Epidemiol. 2017;89:4–11.

21. Vable AM, Canning D, Glymour MM, Kawachi I, Jimenez MP, Subramanian SV. Can social policy influence socioeconomic disparities? Korean War GI Bill eligibility and markers of depression. Ann Epidemiol. 2016;26:129-135.e3.

22. Nguyen QC, Rehkopf DH, Schmidt NM, Osypuk TL. Heterogeneous effects of housing vouchers on the mental health of US adolescents. Am J Public Health. 2016;106:755–62.

23. Howell WG, Wolf PJ, Campbell DE, Peterson PE. School vouchers and academic performance: results from three randomized field trials. J Policy Anal Manage. 2002;21:191–217.

24. White JS, Hamad R, Li X, Basu S, Ohlsson H, Sundquist J, et al. Long-term effects of neighbourhood deprivation on diabetes risk: quasi-experimental evidence from a refugee dispersal policy in Sweden. Lancet Diabetes Endocrinol. 2016;4:517–24.

25. O'Donnell A, Anderson P, Jané-Llopis E, Manthey J, Kaner E, Rehm J. Immediate impact of minimum unit pricing on alcohol purchases in Scotland: controlled interrupted time series analysis for 2015–18. BMJ [Internet]. British Medical Journal Publishing Group; 2019 [cited 2020 Sep 2];366. Available from: https://www.bmj.com/content/366/bmj.l5274

26. Haghpanahan H, Lewsey J, Mackay DF, McIntosh E, Pell J, Jones A, et al. An evaluation of the effects of lowering blood alcohol concentration limits for drivers on the rates of road traffic accidents and alcohol consumption: a natural experiment. The Lancet Elsevier. 2019;393:321–9.

27. Roberto CA, Lawman HG, LeVasseur MT, Mitra N, Peterhans A, Herring B, et al. Association of a beverage tax on sugar-sweetened and artificially sweetened beverages with changes in beverage prices and sales at chain retailers in a large urban setting. JAMA American Medical Association. 2019;321:1799–810.

28. Levitt SD. Using electoral cycles in police hiring to estimate the effects of police on crime: reply. The American Economic Review. American Economic Association; 2002;92:1244–50.

29. Torche F, Sirois C. Restrictive immigration law and birth outcomes of immigrant women. Am J Epidemiol. 2019;188:24–33.

30. Ferdinand AO, Aftab A, Akinlotan MA. Texting-while-driving bans and motor vehicle crash–related emergency department visits in 16 US states: 2007–2014. Am J Public Health. American Public Health Association; 2019;109:748–54.

31. Hamad R, Modrek S, White JS. Paid family leave effects on breastfeeding: a quasi-experimental study of US policies. Am J Public Health. American Public Health Association; 2018;109:164–6.

32. Franckle RL, Thorndike AN, Moran AJ, Hou T, Blue D, Greene JC, et al. Supermarket purchases over the supplemental nutrition assistance program benefit month: a comparison between

participants and nonparticipants. American Journal of Preventive Medicine Elsevier. 2019;57:800–7.

33. Cylus J, Glymour MM, Avendano M. Do generous unemployment benefit programs reduce suicide rates? A state fixed-effect analysis covering 1968–2008. Am J Epidemiol. 2014;180:45–52.

34. Ghiani M, Hawkins SS, Baum CF. Associations between gun laws and suicides. Am J Epidemiol Oxford Academic. 2019;188:1254–61.

35. Lira MC, Xuan Z, Coleman SM, Swahn MH, Heeren TC, Naimi TS. Alcohol policies and alcohol involvement in intimate partner homicide in the U.S. American Journal of Preventive Medicine. Elsevier; 2019;57:172–9.

36. Gertner AK, Rotter JS, Shafer PR. Association between state minimum wages and suicide rates in the U.S. American Journal of Preventive Medicine. Elsevier; 2019;56:648–54.

37. Tessler RA, Mooney SJ, Quistberg DA, Rowhani-Rahbar A, Vavilala MS, Rivara FP. State-level beer excise tax and firearm homicide in adolescents and young adults. Am J Prev Med. 2019;56:708–15.

38. Doyle WR, Skinner BT. Estimating the education-earnings equation using geographic variation. Econ Educ Rev. 2016;53:254–67.

39. Dragone D, Prarolo G, Vanin P, Zanella G. Crime and the legalization of recreational marijuana. J Econ Behav Organ. 2019;159:488–501.

40. Wu G, Boateng FD, Lang X. The spillover effect of recreational marijuana legalization on crime: evidence from neighboring states of Colorado and Washington state. Journal of Drug Issues. SAGE Publications Inc; 2020;0022042620921359.

41. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: Analysis using instrumental variables. JAMA. 1994;272:859–66.

42. Raifman J, Larson E, Barry CL, Siegel M, Ulrich M, Knopov A, et al. State handgun purchase age minimums in the US and adolescent suicide rates: regression discontinuity and difference-in-differences analyses. BMJ. 2020;370:m2436.

43. Delaruelle K, van de Werfhorst H, Bracke P. Do comprehensive school reforms impact the health of early school leavers? Results of a comparative difference-in-difference design. Social Science & Medicine. 2019;239:112542.

44. Bose B, Heymann J. Effects of tuition-free primary education on women's access to family planning and on health decision-making: a cross-national study. Social Science & Medicine. 2019;238:112478.

45. Heflin CM, Ingram SJ, Ziliak JP. The effect of the supplemental nutrition assistance program on mortality. Health Affairs Health Affairs. 2019;38:1807–15.

46. Hughes C. Reexamining the influence of conditional cash transfers on migration from a gendered lens. Demography. 2019;56:1573–605.

47. Lebihan L, Mao Takongmo C-O. Unconditional cash transfers and parental obesity. Soc Sci Med. 2019;224:116–26.

48. Patler C, Hamilton E, Meagher K, Savinar R. Uncertainty about DACA may undermine its positive impact on health for recipients and their children. Health Affairs Health Affairs. 2019;38:738–45.

49. Angrist JD, Keueger AB. Does compulsory school attendance affect schooling and earnings? Q J Econ. 1991;106:979–1014.

50. Angrist JD, Krueger AB. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. Journal of the American Statistical Association. Taylor & Francis; 1992;87:328–36.

51. Angrist JD, Evans WN. Children and their parents' labor supply: evidence from exogenous variation in family size. American Economic Review. 1998;88:450–77.

52. Bronars SG, Grogger J. The Economic consequences of unwed motherhood: using twin births as a natural experiment. The American Economic Review. American Economic Association; 1994;84:1141–56.

53. Fletcher JM. Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts. PLOS ONE. Public Library of Science; 2012;7:e50576.

54. Boardman JD, Fletcher JM. The promise of integrating genetics into policy analysis. Journal of Policy Analysis and Management. [Wiley, Association for Public Policy Analysis and Management]; 2015;34:493–6.

55. Handy S, Cao X, Mokhtarian PL. Self-selection in the relationship between the built environment and walking: empirical evidence from northern California. Journal of the American Planning Association. Routledge; 2006;72:55–74.

56. Cao X, Mokhtarian PL, Handy SL. Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach Transportation. 2007;34:535–56.

57. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. J Clin Epidemiol. 2009;62:1233–41.

58. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. The International Journal of Biostatistics [Internet]. 2007 [cited 2019 Jan 31];3. Available from: https://www.degruyter.com/view/j/ijb.2007.3.1/ijb.2007.3.1.1072/ijb.2007.3.1.1072.xml?format=INT&intcmp=trendmd

59. Kling JR. Incarceration length, employment, and earnings. American Economic Review. 2006;96:863–76.

60. Gifford EJ, Eldred LM, Mccutchan SA, Sloan FA. Prosecution, Conviction, and Deterrence in Child Maltreatment Cases. Criminal Justice and Behavior. SAGE Publications Inc; 2017;44:1262–80.

61.● Tremper C, Thomas S, Wagenaar A. Measuring the law for evaluation research. Evaluation Review. 2010;34:242–66. **An overview of legal epidemiologic methods for identifying and measuring the law for research**.

62. National Institute on Alcohol Abues and Alcoholism. How to measure law for quantitative research: A resource guide [Internet]. [cited 2021 Aug 31]. Available from: https://alcoholpolicy.niaaa.nih.gov/resource/how-to-measure-law-for-quantitative-research-a-resource-guide/18

63. Erickson DJ, Rutledge PC, Lenk KM, Nelson TF, Jones-Webb R, Toomey TL. Patterns of Alcohol Policy Enforcement Activities among Local Law Enforcement Agencies: A Latent Class Analysis. Int J Alcohol Drug Res. 2015;4:103–11.

64. Erickson DJ, Farbakhsh K, Toomey TL, Lenk KM, Jones-Webb R, Nelson TF. Enforcement of alcohol-impaired driving laws in the United States: a national survey of state and local agencies. Traffic Inj Prev. 2015;16:533–9.

65.● Klitzner MD, Thomas S, Schuler J, Hilton M, Mosher J. The new cannabis policy taxonomy on APIS: making sense of the cannabis policy universe. J Primary Prevent. 2017;38:295–314. **An example of expert consensus-driven methods for developing policy taxonomies**.

66. Chapman SA, Spetz J, Lin J, Chan K, Schmidt LA. Capturing heterogeneity in medical marijuana policies: a taxonomy of regulatory regimes across the United States. Subst Use Misuse. 2016;51:1174–84.

67.● Athey S, Imbens GW. The state of applied econometrics: causality and policy evaluation. Journal of Economic Perspectives. 2017;31:3–32. **A useful review of recent advances and key methodological issues in econometric policy studies relevant to epidemiologic research on social policies**.

68. Lin M, Wang Q. Center-based childcare expansion and grandparents' employment and well-being. Social Science & Medicine. 2019;240:112547.

69. Xuan Z, Chaloupka FJ, Blanchette JG, Nguyen TH, Heeren TC, Nelson TF, et al. The relationship between alcohol taxes and binge drinking: evaluating new tax measures incorporating multiple tax and beverage types. Addiction. 2015;110:441–50.

70. Larrimore J. Does a higher income have positive health effects? Using the earned income tax credit to explore the income-health gradient. Milbank Q. 2011;89:694–727.

71. Mair C, Freisthler B, Ponicki WR, Gaidus A. The impacts of marijuana dispensary density and neighborhood ecology on marijuana abuse and dependence. Drug Alcohol Depend. 2015;154:111–6.

72. Freisthler B, Gaidus A, Tam C, Ponicki WR, Gruenewald PJ. From medical to recreational marijuana sales: marijuana outlets and crime in an era of changing marijuana legislation. J Primary Prevent. 2017;38:249–63.

73. Erickson DJ, Lenk KM, Toomey TL, Nelson TF, Jones-Webb R. The alcohol policy environment, enforcement and consumption in the United States. Drug Alcohol Rev. 2016;35:6–12.

74. Rajmil L, Fernández de Sanmamed M-J. Austerity policies and mortality rates in European countries, 2011–2015. Am J Public Health. American Public Health Association; 2019;109:768–70.

75. Bruzelius E, Baum A. The mental health of Hispanic/Latino Americans following national immigration policy changes: United States, 2014–2018. Am J Public Health. American Public Health Association; 2019;109:1786–8.

76. Sivaraman JJ, Ranapurwala SI, Moracco KE, Marshall SW. Association of state firearm legislation with female intimate partner homicide. Am J Prev Med. 2019;56:125–33.

77. Rehkopf DH, Glymour MM, Osypuk TL. The consistency assumption for causal inference in social epidemiology: when a rose is not a rose. Curr Epidemiol Rep. 2016;3:63–71.

78. Pacula RL, Powell D, Heaton P, Sevigny EL. Assessing the effects of medical marijuana laws on marijuana use: the devil is in the details. J Policy Anal Manage. 2015;34:7–31.

79. Shi Y, Liang D. The association between recreational cannabis commercialization and cannabis exposures reported to the US National Poison Data System. Addiction. 2020;115:1890–9.

80. Matthay E, Elser H, Kiang M, Schmidt L, Humphreys K. Evaluation of state cannabis laws and rates of self-harm and assault. JAMA Network Open. 2021;4:e211955.

81. Naimi TS, Blanchette J, Nelson TF, Nguyen T, Oussayef N, Heeren TC, et al. A new scale of the U.S. alcohol policy environment and its relationship to binge drinking. Am J Prev Med. 2014;46:10–6.

82.● Cintron D, Adler N, Dirocco N, Gottlieb L, Glymour M, Hagan E, et al. Heterogenous treatment effects in social policy studies: an assessment of contemporary articles in the health and social sciences. Under review. 2021; **A review of the frequency and methodology of heterogeneous treatment effect evaluation in a contemporary sample of studies on the health effects of social policies**.

83. Hernandez EM, Vuolo M, Frizzell LC, Kelly BC. Moving upstream: the effect of tobacco clean air restrictions on educational inequalities in smoking among young adults. Demography. 2019;56:1693–721.

84.● Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. Epidemiology. 2017;28:553. **A review and illustrative example of concepts and methods for generalizing study results to new target populations—a promising frontier for research on the health effects of social policies**.

85. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. Am J Epidemiol. 2017;186:1010–4.

86. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. Am J Epidemiol. 2010;172:107–15.

87. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. J R Stat Soc A Stat Soc. 2011;174:369–86.

88. Rudolph KE, van der Laan MJ. Robust estimation of encouragement-design intervention effects transported across sites. J R Stat Soc Series B Stat Methodol. 2017;79:1509–25.

89. Ku L, Brantley E, Pillai D. The effects of SNAP work requirements in reducing participation and benefits from 2013 to 2017. Am J Public Health. American Public Health Association; 2019;109:1446–51.

90. Hamad R, Batra A, Karasek D, LeWinn KZ, Bush NR, Davis RL, et al. The impact of the revised WIC food package on maternal nutrition during pregnancy and postpartum. Am J Epidemiol Oxford Academic. 2019;188:1493–502.

91. Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. PNAS National Academy of Sciences. 2016;113:7353–60.

92. List JA, Shaikh AM, Xu Y. Multiple hypothesis testing in experimental economics. Exp Econ. 2019;22:773–93.

93. Chernozhukov V, Demirer M, Duflo E, Fernández-Val I. Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India [Internet]. National Bureau of Economic Research; 2018 Jun. Report No.: 24678. Available from: https://www.nber.org/papers/w24678

94. Athey S, Wager S. Efficient Policy Learning. arXiv:170202896 [cs, econ, math, stat] [Internet]. 2017 [cited 2019 Feb 13]; Available from: http://arxiv.org/abs/1702.02896

95. Zhou Z, Athey S, Wager S. Offline Multi-Action Policy Learning: Generalization and Optimization. 2018 [cited 2021 Dec 7]; Available from: https://arxiv.org/abs/1810.04778v2

96. Starks MA, Sanders GD, Coeytaux RR, Riley IL, Ii LRJ, Brooks AM, et al. Assessing heterogeneity of treatment effect analyses in health-related cluster randomized trials: a systematic review. PLOS ONE. Public Library of Science; 2019;14:e0219894.

97. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ. British Medical Journal Publishing Group; 2012;344:e1553.

98. Kasenda B, Schandelmaier S, Sun X, Elm E von, You J, Blümle A, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. BMJ. British Medical Journal Publishing Group; 2014;349:g4539.

99. Fan J, Song F, Bachmann MO. Justification and reporting of subgroup analyses were lacking or inadequate in randomized controlled trials. J Clin Epidemiol. 2019;108:17–25.

100. Fernandez y Garcia E, Nguyen H, Duan N, Gabler NB, Kravitz RL. Assessing heterogeneity of treatment effects: are authors misinterpreting their results? Health Services Research. 2010;45:283–301.

101. Doleac JL, Hansen B. Does "Ban the Box" Help or Hurt Low-Skilled Workers? Statistical Discrimination and Employment Outcomes When Criminal Histories are Hidden [Internet]. National Bureau of Economic Research; 2016 Jul. Report No.: 22469. Available from: https://www.nber.org/papers/w22469

102. Kaufman EJ, Morrison CN, Branas CC, Wiebe DJ. State firearm laws and interstate firearm deaths from homicide and suicide in the United States: a cross-sectional analysis of data by county. JAMA Intern Med. 2018;178:692–700.

103. Matthay EC, Galin J, Rudolph KE, Farkas K, Wintemute GJ, Ahern J. In-state and interstate associations between gun shows and firearm deaths and injuries: a quasi-experimental study. Ann Intern Med. 2017;167:837.

104. Schneider SK, Buka SL, Dash K, Winickoff JP, O'Donnell L. Community reductions in youth smoking after raising the minimum tobacco sales age to 21. Tobacco Control BMJ Publishing Group Ltd. 2016;25:355–9.

105. Falbe J, Thompson HR, Becker CM, Rojas N, McCulloch CE, Madsen KA. Impact of the Berkeley excise tax on sugar-sweetened beverage consumption. Am J Public Health. American Public Health Association; 2016;106:1865–71.

106. Tabb LP, Ballester L, Grubesic TH. The spatio-temporal relationship between alcohol outlets and violence before and after privatization: a natural experiment, Seattle, Wa 2010–2013. Spatial and Spatio-temporal Epidemiology. 2016;19:115–24.

107. Matthay E, Mousli L, Ponicki W, Glymour M, Apollonio D, Schmidt L, et al. Local policies as instruments for health equity? A spatiotemporal analysis of the association of California city and county cannabis policies with cannabis dispensary densities. Under review. 2021;

108. Keyes KM, Shev A, Tracy M, Cerdá M. Assessing the impact of alcohol taxation on rates of violent victimization in a large urban area: an agent-based modeling approach. Addiction. 2019;114:236–47.

109. Cerdá M, Jalali MS, Hamilton AD, DiGennaro C, Hyder A, Santaella-Tenorio J, et al. A systematic review of simulation models to track and address the opioid crisis. Epidemiol Rev. 2021;mxab013.

110. Tchetgen EJT, VanderWeele TJ. On causal inference in the presence of interference. Stat Methods Med Res. 2012;21:55–75.

111. Muñoz ID, van der Laan M. Population intervention causal effects based on stochastic interventions. Biometrics. 2012;68:541–9.

112. Goldsmith-Pinkham P, Imbens GW. Social networks and the identification of peer effects. Journal of Business & Economic Statistics. Taylor & Francis; 2013;31:253–64.

113. Bramoullé Y, Djebbari H, Fortin B. Identification of peer effects through social networks. Journal of Econometrics. 2009;150:41–55.

114.● Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target validity and the hierarchy of study designs. Am J Epidemiol. 2019;188:438–43. A useful perspective on integrating concepts of internal and external validity. **The arguments presented support the application of transportability estimators to observational and quasi-experimental studies of social policies**.

115.● Matthay EC, Hagan E, Gottlieb LM, Tan ML, Vlahov D, Adler N, et al. Powering population health research: considerations for plausible and actionable effect sizes. SSM - Population Health. 2021;14:100789. **Discusses important considerations for informing calculations of power, sample size, and minimum detectable effect for studies on the health effects of social policies**.

116. Rose G. Sick individuals and sick populations. Int J Epidemiol Oxford Academic. 2001;30:427–32.

117.● Courtin E, Kim S, Song S, Yu W, Muennig P. Can social policies improve health? A systematic review and meta-analysis of 38 randomized trials. The Milbank Quarterly. 2020;98:297–371. **Systematic review of evidence on health effects from randomized social policy experiments**.

118. Langa KM, Plassman BL, Wallace RB, Herzog AR, Heeringa SG, Ofstedal MB, et al. The aging, demographics, and memory study: study design and methods. NED. 2005;25:181–91.

119. Gottlieb L, Tobey R, Cantor J, Hessler D, Adler NE. Integrating social and medical data to improve population health:

120. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLOS Medicine. 2015;12:e1001779.

121. Black B, Hollingsworth A, Nunes L, Simon K. Simulated Power Analyses for Observational Studies: An Application to the Affordable Care Act Medicaid Expansion [Internet]. National Bureau of Economic Research; 2019 Feb. Report No.: 25568. Available from: https://www.nber.org/papers/w25568

122. Ioannidis JPA, Stanley TD, Doucouliagos H. The power of bias in economics research. Econ J. 2017;127:F236–65.

123. Smart R, Morral AR, Smucker S, Cherney S, Schell TL, Peterson S, et al. The Science of Gun Policy: A Critical Synthesis of Research Evidence on the Effects of Gun Policies in the United States, Second Edition [Internet]. RAND Corporation; 2020 Apr. Available from: https://www.rand.org/pubs/research_reports/RR2088-1.html

124. Smith N, Martins SS, Kim J, Rivera-Aguirre A, Fink DS, Castillo-Carniglia A, et al. A typology of prescription drug monitoring programs: a latent transition analysis of the evolution of programs from 1999 to 2016. Addiction. 2019;114:248–58.

125. Friedman SR, Krawczyk N, Perlman DC, Mateu-Gelabert P, Ompad DC, Hamilton L, et al. The opioid/overdose crisis as a dialectics of pain, despair, and one-sided struggle. Front Public Health. 2020;8:540423.

126. Hernán MA, Robins J. G-methods for time-varying treatments. Causal inference: What if. Charleston, SC: BN Publishing; 2020. p. 247–54.

127. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. BMJ. British Medical Journal Publishing Group; 2017;359:j4587.

128. Vansteelandt S, Sjolander A. Revisiting g-estimation of the effect of a time-varying exposure subject to time-varying confounding. Epidemiologic Methods De Gruyter. 2016;5:37–56.

129.● Clare PJ, Dobbins TA, Mattick RP. Causal models adjusting for time-varying confounding—a systematic review of the literature. Int J Epidemiol. 2019;48:254–65. **Reviews the common methods for addressing time-varying confounding—an underaddressed issue in research on the health effects of social policies**.

130. Orr LL. Social Experiments: Evaluating Public Programs With Experimental Methods. SAGE; 1999.

131. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol. 2006;6:54.

132. Handley MA, Schillinger D, Shiboski S. Quasi-experimental designs in practice-based research settings: design and implementation considerations. J Am Board Fam Med. American Board of Family Medicine; 2011;24:589–96.

133.● Wintemute GJ, Beckett L, Kass PH, Tancredi D, Studdert D, Pierce G, et al. Evaluation of California's armed and prohibited persons system: study protocol for a cluster-randomised trial. Injury Prevention. BMJ Publishing Group Ltd; 2017;23:358–358. **A striking example of how a randomized trial can be embedded rollout of a large-scale policy**.

134. LaLonde RJ. Evaluating the econometric evaluations of training programs with experimental data. Am Econ Rev. 1986;76:604–20.

135. Dehejia RH, Wahba S. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. J Am Stat Assoc. 1999;94:1053–62.

136.● Glynn AN, Kashin K. Front-door versus back-door adjustment with unmeasured confounding: bias formulas for front-door

and hybrid adjustments with application to a job training program. Journal of the American Statistical Association. Taylor & Francis; 2018;113:1040–9. **Notable study demonstrating the robustness of the front door criterion for causal identification.**

137. Bellemare MF, Bloem JR, Wexler N. The Paper of How: Estimating Treatment Effects Using the Front-Door Criterion [Internet]. 2020 Sep p. 47. Available from: http://marcfbellemare.com/wordpress/wp-content/uploads/2020/09/BellemareBloemWexlerFDCSeptember2020.pdf

138. Fulcher IR, Shpitser I, Marealle S, Tchetgen Tchetgen EJ. Robust inference on population indirect causal effects: the generalized front door criterion. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2020;82:199–214.

139. Díaz I, Williams N, Hoffman KL, Schenck EJ. Nonparametric causal effects based on longitudinal modified treatment policies. Journal of the American Statistical Association. Taylor & Francis; 2021;0:1–16.

140.● Rudolph K, Gimbrone C, Matthay E, Diaz I, Davis C, Keyes KM, et al. When effects cannot be estimated: redefining estimands to understand the effects of naloxone access laws. Under review. 2021; **An illustration of estimation of a novel policy effect parameter based on shifts in the timing of policy adoption to address practical positivity violations**.

141.● Roth J, Sant'Anna PHC, Bilinski A, Poe J. What's trending in differences-in-differences? A synthesis of the recent econometrics literature. 2022 Jan. **A review of recent advances in the econometrics literature on differences-in-differences, including important biases in the two-way fixed effects design and newly developed estimators that address these biases.**

142. Angrist J, Pischke J-S. 3.1 Regression Fundamentals. Mostly Harmless Econometrics. Princeton, New Jersey: Princeton University Press; 2008.

143. de Cuba SAE, Bovell-Ammon AR, Cook JT, Coleman SM, Black MM, Chilton MM, et al. SNAP, Young children's health, and family food security and healthcare access. American Journal of Preventive Medicine Elsevier. 2019;57:525–32.

144. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci. 2010;25:1–21.

145. Shahidi FV, Muntaner C, Shankardass K, Quiñonez C, Siddiqi A. The effect of unemployment benefits on health: a propensity score analysis. Soc Sci Med. 2019;226:198–206.

146.● Doudchenko N, Imbens G. Balancing, regression, difference-in-differences, and synthetic control methods: a synthesis. NBER Working Paper No 22791 arXiv. 2016; **A useful elaboration of the relationship between different causal inference methods commonly used in econometrics**.

147. Matthay EC, Farkas K, Rudolph KE, Zimmerman S, Barragan M, Goin DE, et al. Firearm and nonfirearm violence after operation peacemaker fellowship in Richmond, California, 1996–2016. Am J Public Health. 2019;109:1605–11.

148. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000;29:722–9.

149. Chen X, Wang T, Busch SH. Does money relieve depression? Evidence from social pension expansions in China. Soc Sci Med. 2019;220:411–20.

150. Bor J, Moscoe E, Mutevedzi P, Newell M-L, Bärnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology. 2014;25:729–37.

151. Stacey N, Mudara C, Ng SW, van Walbeek C, Hofman K, Edoka I. Sugar-based beverage taxes and beverage prices: evidence from South Africa's Health Promotion Levy. Social Science & Medicine. 2019;238:112465.

152. Lin B-H, Guthrie JF, Smith TA. Dietary guidance and new school meal standards: schoolchildren's whole grain consumption over 1994–2014. American Journal of Preventive Medicine Elsevier. 2019;57:57–67.

153. Evans WN, Kroeger S, Palmer C, Pohl E. Housing and urban development–veterans affairs supportive housing vouchers and veterans' homelessness, 2007–2017. Am J Public Health. American Public Health Association; 2019;109:1440–5.

154. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. American College of Physicians; 2017;167:268–74.

155. Lash T, Fink A, Fox MP. A guide to implementing quantitative bias analysis. Applying Quantitative Bias Analysis to Epidemiologic Data. New York, NY: Springer; 2009. p. 13–32.

156. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. Int J Epidemiol. 2014;43:1969–85.

157.● Banack HR, Hayes-Larson E, Mayeda ER. Monte Carlo simulation approaches for quantitative bias analysis: a tutorial. Epidemiologic Reviews. 2022;43:106–17. **Reviews key concepts for quantitative bias analysis—an underutilized tool to enhance the validity of research on the health effects of social policies.**

158. Lash TL, Ahern TP, Collin LJ, Fox MP, MacLehose RF. Bias analysis gone bad. Am J Epidemiol. 2021;190:1604–12.

159. Greenland S. Invited commentary: dealing with the inevitable deficiencies of bias analysis—and all analyses. Am J Epidemiol. 2021;190:1617–21.

160. Lipsitch M, Tchetgen ET, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21:383–8.

161. Castillo-Carniglia A, Kagawa RMC, Cerdá M, Crifasi CK, Vernick JS, Webster DW, et al. California's comprehensive background check and misdemeanor violence prohibition policies and firearm mortality. Ann Epidemiol. 2019;30:50–6.

162.● Cerdá M, Keyes KM. Systems modeling to advance the promise of data science in epidemiology. American Journal of Epidemiology. 2019;188:862–5. **Discusses the contributions, challenges, and future directions for complex systems modeling with applications to policy studies. Reviews the common methods for addressing time-varying confounding—an underaddressed issue in research on the health effects of social policies**.

163. Evidence for Action. Using Complex Systems Modeling for Social Good: An Interview with George Kaplan | Evidence for Action [Internet]. 2020 [cited 2021 Dec 14]. Available from: https://www.evidenceforaction.org/blog-posts/using-complex-systems-modeling-social-good-interview-george-kaplan

164. Herlands W, McFowland E, Wilson A, Neill D. Automated local regression discontinuity design discovery. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018. p. 1512–20. 165.   l   a Bastide-van GS, Stolk RP, van den Heuvel ER, Fidler V. Causal inference algorithms can be useful in life course epidemiology. J Clin Epidemiol. 2014;67:190–8.

An applied example of a study analyzing the health effects of social policies

166. Torche F, Sirois C. Restrictive Immigration Law and Birth Outcomes of Immigrant Women. American Journal of Epidemiology. 2019;188:24–33.

167. Lin M, Wang Q. Center-based childcare expansion and grandparents' employment and well-being. Social Science & Medicine. 2019;240:112547.

168. Hamad R, Batra A, Karasek D, LeWinn KZ, Bush NR, Davis RL, et al. The Impact of the Revised WIC Food Package on

Maternal Nutrition During Pregnancy and Postpartum. Am J Epidemiol. Oxford Academic; 2019;188:1493–502.

169. Smith N, Martins SS, Kim J, Rivera-Aguirre A, Fink DS, Castillo-Carniglia A, et al. A typology of prescription drug monitoring programs: a latent transition analysis of the evolution of programs from 1999 to 2016. Addiction. 2019;114:248–58.

170. Roberto CA, Lawman HG, LeVasseur MT, Mitra N, Peterhans A, Herring B, et al. Association of a Beverage Tax on Sugar-Sweetened and Artificially Sweetened Beverages With Changes in Beverage Prices and Sales at Chain Retailers in a Large Urban Setting. JAMA. American Medical Association; 2019;321:1799–810.

171. Matthay EC, Farkas K, Rudolph KE, Zimmerman S, Barragan M, Goin DE, et al. Firearm and Nonfirearm Violence After Operation Peacemaker Fellowship in Richmond, California, 1996–2016. Am J Public Health. 2019;109:1605–11.

172. Haghpanahan H, Lewsey J, Mackay DF, McIntosh E, Pell J, Jones A, et al. An evaluation of the effects of lowering blood alcohol concentration limits for drivers on the rates of road traffic accidents and alcohol consumption: a natural experiment. The Lancet. Elsevier; 2019;393:321–9.

173. Keyes KM, Shev A, Tracy M, Cerdá M. Assessing the impact of alcohol taxation on rates of violent victimization in a large urban area: an agent-based modeling approach. Addiction. 2019;114:236–47.

174. Erickson DJ, Farbakhsh K, Toomey TL, Lenk KM, Jones-Webb R, Nelson TF. Enforcement of alcohol-impaired driving laws in the United States: a national survey of state and local agencies. Traffic Inj Prev. 2015;16:533–9.