Psychological Trauma: Theory, Research, Practice, and Policy

Opportunities and Challenges in Using Instrumental Variables to Study Causal Effects in Nonrandomized Stress and Trauma Research

Ellicott C. Matthay, Meghan L. Smith, M. Maria Glymour, Justin S. White, and Jaimie L. Gradus

Online First Publication, October 13, 2022. http://dx.doi.org/10.1037/tra0001370

AMERICAN PSYCHOLOGICAL ASSOCIATION
TRAUMA PSYCHOLOGY DIVISION 56 AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Trauma:
Theory, Research, Practice, and Policy

ISSN: 1942-9681
https://doi.org/10.1037/tra0001370

Opportunities and Challenges in Using Instrumental Variables to Study Causal Effects in Nonrandomized Stress and Trauma Research

Ellicott C. Matthay[1], Meghan L. Smith[2], M. Maria Glymour[3], Justin S. White[4], and Jaimie L. Gradus[2]
[1] Center for Opioid Epidemiology and Policy, Division of Epidemiology, Department of Population Health, New York University Grossman School of Medicine
[2] Department of Epidemiology, Boston University School of Public Health
[3] Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco
[4] Philip R. Lee Institute for Health Policy Studies, School of Medicine, University of California, San Francisco

Objective: Researchers are often interested in assessing the causal effect of an exposure on an outcome when randomization is not ethical or feasible. Estimating causal effects by controlling for confounders can be unconvincing because important potential confounders remain unmeasured. Study designs leveraging instrumental variables (IVs) offer alternatives to confounder–control methods but are rarely used in stress and trauma research. ***Method:*** We review the conceptual foundations and implementation of IV methods. We discuss strengths and limitations of IV approaches, contrasting with confounder–control methods, and illustrate the relevance of IVs for stress and trauma research. ***Results:*** IV approaches leverage an external or exogenous source of variation in the exposure. Instruments are variables that meet three conditions: relevance (variation in the IV is associated with variation in the chance of exposure), exclusion (the IV only affects the outcome through the exposure), and exchangeability (no unmeasured confounding of the IV–outcome relationship). Interpreting estimates from IV analyses requires an additional assumption, such as monotonicity (the instrument does not change the chance of exposure in different directions for any two individuals). Valid IVs circumvent the need to correctly identify, measure, and control for all confounders of the exposure–outcome relationship. The primary challenge is identifying a valid instrument. ***Conclusions:*** IV approaches have strengths and weaknesses compared with confounder–control approaches. IVs offers a promising complementary study design to improve evidence about the causal effects of exposures on outcomes relevant to stress and trauma. Collaboration with scientists who are experienced with identifying and analyzing IVs will support this work.

Clinical Impact Statement

Estimating the causal effects of stressful or traumatic experiences on later health is important but difficult because randomizing stressful or traumatic experiences or trauma-informed interventions may not be ethical. Instead, stress and trauma researchers typically account for measured confounders using regression, matching, or stratification methods. Study designs leveraging instrumental variables (IV) methods offer a valuable alternative approach, relying on alternative assumptions. In observational stress and trauma research, complementing confounder–control studies with IV-based methods is likely to strengthen the body of evidence on causal effects and help inform future interventions.

Keywords: causal inference, confounding, instrumental variable, stress, trauma

Ellicott C. Matthay https://orcid.org/0000-0003-4535-8252

This work was supported by Award K99AA028256 from the National Institute on Alcohol Abuse and Alcoholism, the Evidence for Action program of the Robert Wood Johnson Foundation, and Award R01MH110453 from the National Institute of Mental Health. The funding source had no role in the study design interpretation, analysis, writing, or decision to submit for publication. The authors thank Sonja Swanson for her helpful feedback on this article.

Correspondence concerning this article should be addressed to Ellicott C. Matthay, Center for Opioid Epidemiology and Policy, Division of Epidemiology, Department of Population Health, New York University Grossman School of Medicine, 180 Madison Avenue, New York, New York 10016, United States. Email: ellicott.matthay@nyulangone.org

In stress and trauma research, investigators often conduct statistical analyses to document the relationship between an exposure and an outcome. We use the term “exposure” broadly to refer to any treatment, experience, event, or condition that may influence an outcome of interest. Documenting associations between an exposure and outcome typically has one of two goals: (a) to evaluate whether the exposure may be a *cause* of the outcome, or (b) to evaluate whether the exposure is *predictive* of the outcome, regardless of whether the exposure actually causes the outcome. Identifying whether a traumatic experience *predicts* an outcome can be useful to determine whether an exposed group needs services. For example, documenting the extent to which rural residence is associated with suicide can ensure that adequate resources are directed to suicide prevention in rural areas. Even if rural residence does not *cause* suicide, the association informs service allocation. In contrast, estimating *causal* effects is essential to evaluate whether a specific exposure *caused* an outcome and thus whether *intervening* on that exposure is likely to change the outcome. Whereas many estimates in the stress and trauma literature are characterized as associational, researchers are often implicitly interested in estimating causal effects. For example, researchers may seek to understand whether military service *causes* posttraumatic stress disorder (PTSD), so that clinicians can understand whether military service is a reason PTSD arises, and decisionmakers can understand how *changing* deployments will affect the number of veterans with PTSD (Angrist et al., 2010; Conley & Heerwig, 2012; Gade & Wenger, 2011; Johnston et al., 2016; Lyk-Jensen et al., 2016).

Confounding: A Primary Challenge for Causal Inference

A primary challenge in estimating causal effects is conceptualizing and adjusting for bias owing to confounding (Matthay et al., 2019). Confounding arises when there are differences between exposed and unexposed individuals with regard to variables that also affect the outcome of interest. For example, when estimating the effect of military service on PTSD, a possible study design would compare the health outcomes of people with and without military service experience. However, military service involves both eligibility and interest in participation (even in the context of a draft, some people volunteered). Military veterans may have different physical, cognitive, personality, or behavioral profiles than nonparticipants even prior to enlisting, and if these profiles influence PTSD risk, military service and PTSD will be spuriously associated. Left unaddressed, these unbalanced factors (i.e., confounders) can induce confounding bias in estimates of the causal effect of military service. Thus, the measured association between the exposure and the outcome may differ from the true causal effect of the exposure on the outcome due to confounding.

Randomized trials can offer strong evidence of the causal effect of an exposure on an outcome, because randomization ensures that on average any potential confounders, whether measured or unmeasured, are balanced between the exposed and unexposed groups and thus will not lead to confounding bias. However, there are many situations in which an exposure (war, natural disasters, homelessness, violent victimization, or other traumatic and stressful life experiences) is not ethical or feasible to randomize. Randomizing access to trauma-focused interventions can also be fraught. Research questions often arise after an exposure has occurred or outside of a randomized trial, and because of restrictive eligibility criteria, randomized trials have limited generalizability. Thus, many questions pertinent to stress and trauma can only be examined in observational research.

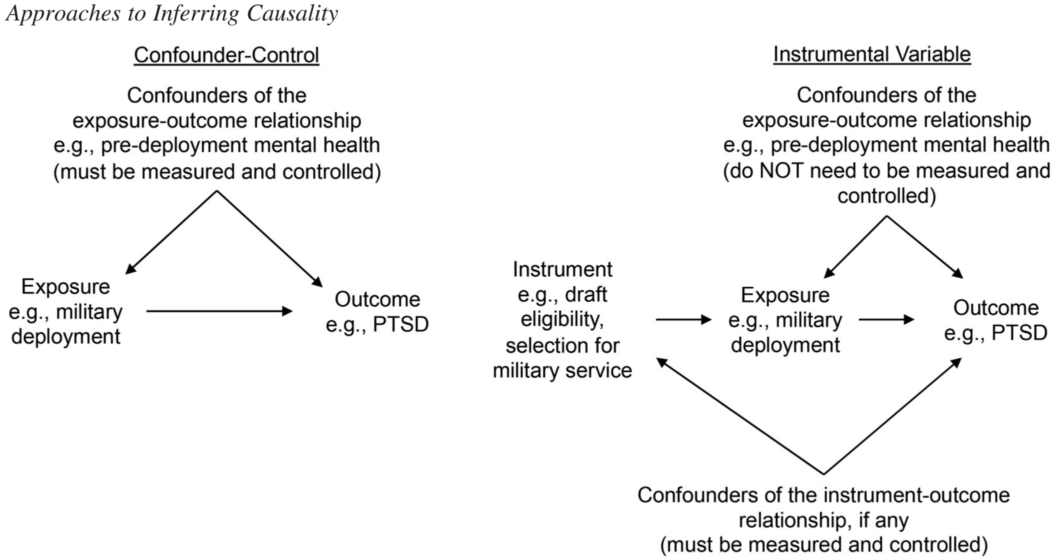
Two Observational Approaches to Addressing Confounding

In observational research, we distinguish between two main strategies for addressing confounding to estimate causal effects: confounder–control and instrumental variables (see Figure 1; Matthay et al., 2019; Pearl, 2009). Confounder–control approaches involve correctly identifying, measuring, and adjusting for the confounding variables. The vast majority of nonexperimental stress and trauma research uses this approach, and some common examples of these statistical approaches are multiple regression (Fernandez et al., 2017), propensity score matching (Ebrahimi et al., 2021), propensity score weighting (Scherrer et al., 2020), stratified analyses, and structural equations modeling (Syed Sheriff et al., 2019). However, confounder–control studies can fall short of estimating causal effects because of incomplete adjustment (i.e., not all confounders are observed, adequately measured, or adjusted) or inappropriate adjustment (i.e., adjustment for factor that are not confounders; Jiang et al., 2021).

Instrumental variable (IV) approaches leverage an external or exogenous source of variation in the exposure, known as an

Figure 1

Causal Diagrams (Directed Acyclic Graphs) Contrasting Confounder–Control and Instrumental Variable Approaches to Inferring Causality



instrument. Examples of exogenous variation include lotteries, arbitrary changes or thresholds used by a program or policy, or other "accidents" of time or space. For example, researchers have argued that among individuals surveyed through the ongoing Behavioral Risk Factor Surveillance System, individuals who were interviewed immediately before versus immediately after the terrorist attacks of September 11, 2001, were very similar except with respect to their short-term stress levels (Pesko & Baum, 2016). This means that differences in the interview date (days before or after the attack) could be used as an instrument for short-term stress levels to study how stress affects smoking.

This IV-based approach to estimating the causal effect of short-term stress on smoking can be compelling because there are many confounders of the stress-smoking relationship (e.g., prior psychopathology). Accurately identifying, measuring, and adjusting for these confounders can be challenging. In contrast, IV approaches do not require to the investigator to correctly adjust for all confounders of the exposure–outcome relationship, but instead rely on an alternative set of assumptions described in the next section.

The Potential of Instrumental Variable Methods

IV-based designs have a long history of use in the social sciences, and various types of IV methods predominate contemporary applied econometrics research (Angrist & Krueger, 2001; Athey & Imbens, 2017). In stress and trauma research, applications of IV approaches are rare, but the field may benefit from diversifying the set of available causal inference tools. Both confounder–control approaches and IV-based methods rely on assumptions; the preferred approach depends on the context, and no one approach is universally preferable. Because neither set of assumptions can be tested, conducting studies that vary the required assumptions may contribute a stronger body of evidence than relying on any one approach alone. Barriers to the use of IV methods may include lack of exposure to IV concepts and limited experience with how the assumptions required by IV methods can be met in the setting of stress and trauma research. This article aims to fill these gaps.

The present article offers an introduction to the conceptual approach, key assumptions, and implementation of IV methods. Drawing on examples from the stress and trauma literature, we discuss the strengths, weaknesses, and tradeoffs when comparing IVs and confounder–control approaches. We aim to foster understanding of alternative causal inference methods, help investigators identify opportunities for strengthening causal inferences in the absence of randomization, and facilitate the selection of methods best suited to each unique scientific question.

Key Concepts of Instrumental Variables

Defining an Instrument

Instruments are variables, factors, or influences associated with differences in the likelihood of exposure or treatment between people or units that are otherwise similar. Instruments are often called "exogenous" factors because they can be thought of as an external influence on a system. Instruments are described as "good-as-random" or "quasi-random" because valid instruments imply that people who are exposed versus unexposed are, on average, very similar, as would be true in a randomized trial.

For an IVs analysis to validly evaluate whether an exposure causes an outcome (unbiased by confounding), three conditions or assumptions must be met. Formal descriptions of these conditions based on the potential outcomes framework are presented elsewhere (Angrist et al., 1996; Pearl, 2009). In Table 1 we present formal definitions of the conditions based on the causal framework of directed acyclic graphs (DAGs) from Pearl (2009). For an introduction to DAGs, we direct the reader to Austin et al. (2019). We also describe these conditions informally to help readers recognize potential IVs. The first condition, *relevance*, states that different values of the IV result in differences in the chance of exposure.[1] The second condition, *exclusion*, states that the instrument has no influence on the outcome except through the exposure. The third condition, *exchangeability* (i.e., independence or exogeneity), states that there are no unmeasured confounders of the instrument–outcome relationship (i.e., the instrument does not share unmeasured causes with the outcome). When the instrument is random assignment, as in a randomized trial, exchangeability is especially plausible. Each of these conditions can be met as-is or by conditioning on measured covariates. Relevance, exclusion, and exchangeability are sufficient to evaluate causality, but to interpret estimates from IV analyses, researchers must rely on a fourth condition, addressed in the section on estimation and interpretation.

We illustrate the first three conditions with an example. Suppose we aim to estimate the effect of experiencing damage to one's home as a result of a tsunami on cognitive decline (Hikichi et al., 2016), and we propose that one instrument could be the distance an individual lived from the coast when the tsunami occurred. Relevance implies that distance from the coast is related to the amount of damage to one's home; this condition can be directly tested. Exclusion means that the only reason that distance to the coast is relevant to cognitive decline is because proximity to the coast led to greater housing damage. This would be violated, for example, if living close to the coast led people to adopt distinctive dietary patterns or leisure time activities that influenced cognition. Exchangeability means that there are no unmeasured confounders (or common causes) of distance from the coast and cognitive decline. For example, if wealthier people lived nearer to the coast and also had better access to resources that prevent cognitive decline, this pattern would violate exchangeability.

Researchers applying an IV approach are responsible for evaluating the plausibility of the conditions for their planned application. Relevance can be tested empirically by measuring the association of the instrument with the exposure. Exclusion and exchangeability cannot be proven or tested directly, although researchers have developed several falsification tests to evaluate the plausibility of assumptions under specific conditions (Labrecque & Swanson, 2018; Pizer, 2016). The likelihood that they are met in a given application is judged based on substantive knowledge, prior research, or expert opinion.

We contrast IV conditions with the assumptions required for confounder–control (see Table 1): For a confounder–control

[1] The relevance condition can also be fulfilled when the IV and exposure are associated but the IV does not *cause* the exposure (i.e., if the association arises due to a third variable influencing both the IV and the exposure), as long as that the third variable itself fulfills the relevance, exclusion, and exchangeability criteria. This modification is not essential to understand IVs but is invoked in many applications.

Table 1

Formal (Directed Acyclic Graph) and Informal Conditions Required for Evaluating Causality Using Confounder–Control Versus Instrument-Based Approaches

Approach	Condition	Formal conditions (Pearl, 2009)	Informal conditions
Confounder–control	Exchangeability	There exists a set of measured variables for which (a) no variable in the set is a descendant of the exposure, and (b) the set blocks every path between the exposure and the outcome that contains an arrow into the exposure (i.e., there are no unblocked backdoor paths from the exposure to the outcome).	A set of covariates sufficient to control confounding between the exposure and outcome have been correctly identified, measured, and statistically controlled (i.e., there is no unmeasured confounding of the exposure–outcome relationship).
Instrumental variable	Relevance	The instrument is not independent of the exposure (e.g., because the instrument causally affects the exposure or because the instrument shares a common cause with the exposure).	The instrument is associated with the exposure.
	Exclusion	Every directed path connecting the instrument to the outcome contains an arrow pointing into the exposure (i.e., there are no paths via which the instrument influences the outcome that do not pass through the exposure).	The instrument does not affect the outcome except through its potential effect on the exposure.
	Exchangeability	There exists a set of measured variables for which (a) no variable in the set is a descendant of the instrument, and (b) the set blocks every path between the instrument and the outcome that contains an arrow into the instrument (i.e., there are no unblocked backdoor paths from the instrument to the outcome such as shared causes of the instrument and outcome).	The instrument and the outcome do not share unmeasured causes (i.e., there is no unmeasured confounding of the instrument–outcome relationship).
	Monotonicity	The effect of the instrument on the probability of exposure is monotonic (i.e., the instrument always increases or has no impact on the likelihood of exposure, or the instrument always decreases or has no impact on the likelihood of exposure) for all units.	The instrument does not change the likelihood of exposure in different directions for any two individuals.

Note. The table provides formal directed acyclic graph-based definitions of the conditions or assumptions required for evaluating causality (from Pearl, 2009) as well as informal definitions. The formal assumptions can be stated in different ways and we refer the reader to Angrist et al. (1996), Hernán and Robins (2020), Pearl (2009), and Glymour and Swanson (2021) for different variations. The first three conditions define the IV. There are several options for the fourth IV assumption to support causal interpretation of estimates from IV analyses. The most commonly adopted assumption is monotonicity, but others may be preferable in different contexts (see next section). The combination of relevance, exclusion, exchangeability, and monotonicity results in estimates of the local average treatment effect (LATE)—that is, the causal effect of the variation in the exposure that is induced by the instrument on the outcome. Assuming homogeneity of treatment effects across population subgroups results in estimates of the population average treatment effect (PATE)—that is, the average difference in the outcome if everyone in the population were exposed versus if no one in the population were exposed. Without the monotonicity assumption, the quantity estimated by an IV analysis cannot be interpreted as an average causal effect (Angrist et al., 1996), although one might be able to put bounds on the LATE or PATE, but it does not give a point estimate for the magnitude of effect. For more detail, see section "Additional resources for learning more about instrumental variable methods."

approach to deliver a valid estimate of the causal effect of an exposure on an outcome, the investigator must identify, measure, and statistically control for a set of covariates that is sufficient to eliminate confounding (i.e., exchangeability). In our example of estimating the effect of tsunami-induced housing damage on cognitive aging, the exchangeability condition for confounder–control means that all of the factors that (a) precede the tsunami but influence the degree of housing damage and (b) also affect cognitive decline are fully measured and accounted for in the statistical analysis. In practice, there are numerous plausible confounders—wealth, capacity of the resident to maintain their house, resources the resident deployed to secure their house from flooding. As with IVs, confounder–control assumptions cannot be tested and have to be judged substantively.

One intuition for why IV analyses circumvent exposure–outcome confounding is to recognize that although some of the variation in the exposure is related to the confounders, the variation in the exposure that is induced by the instrument is independent of the confounders. IV analyses quantify the effect of interest using just the variation in the exposure that is unrelated to confounders and applies only to people for whom the exposure was independent of the exposure–outcome confounders. Thus, the IV is used to estimate the causal effect for the hypothetical subgroup whose exposure value is affected by the IV. We return to this point when we discuss the local average treatment effect.

Sources of Instruments

Despite the challenge of identifying valid instruments, many instruments in the existing literature across disciplines offer promise for stress and trauma research (see Table 2). Lotteries (e.g., wartime draft lotteries and lotteries for housing vouchers or other resources) are common sources of instruments (Palmer et al., 2019; Singhal, 2019; White et al., 2016). Arbitrary discontinuities or cutoffs can also be instruments. For example, when a new program is implemented on an arbitrary date, or an unexpected event occurs, the days immediately before and after that date are often very similar except for presence of the event (Pesko & Baum, 2016; Torche, 2018; Tsai & Venkataramani, 2015). Arbitrary residential moves, such as those induced by compulsory relocation of military families, have been used to study the causal effects of place-based exposures on health (Lleras-Muney, 2010). Biological chance, such as the genetic variants someone inherits (Bountress et al., 2021; Gu et al., 2021; Sumner et al., 2020) or the sex of a child (Angrist & Evans, 1998; Bronars & Grogger, 1994) have been used as instruments. Using genetic variants as the instrument (i.e., Mendelian Randomization) is useful because genetic variants can influence likelihood of behaviors (e.g., alcohol consumption) or conditions (e.g., obesity). Genetic variants can thus be leveraged to understand the causal effects of those behaviors or

Table 2
Potential Sources of Instruments in Stress and Trauma Research

Instrument type	Examples
Physical distance from place of residence to traumatic exposure	Instrument: Distance from district of residence to location where Vietnam war bombing was heaviest Exposure: Early-life exposure to war Outcomes: Severe mental distress (Singhal, 2019); disability (Palmer et al., 2019) Instrument: Distance from coast Exposure: Housing damage related to earthquake and tsunami Outcome: Cognitive decline (Hikichi et al., 2016)
Lotteries and random assignment	Instrument: Wartime draft lottery Exposure: Military service Outcomes: Mortality (Conley & Heerwig, 2012; Johnston et al., 2016) Instrument: Housing vouchers Exposure: Neighborhood poverty and housing discrimination Outcome: Psychological distress and major depressive disorder (Osypuk et al., 2019) Instrument: Refugees' assignment of residential address by government Exposure: Neighborhood deprivation Outcome: Diagnosis of type 2 diabetes (White et al., 2016)
Discontinuities based on dates of policy changes or disasters	Instrument: Date relative to implementation of universal primary education policy Exposure: Educational attainment Outcome: HIV-related stigma (Tsai & Venkataramani, 2015) Instrument: Days to/since terrorist attacks of September 11, 2001 Exposure: Short-term stress Outcome: Cigarette smoking (Pesko & Baum, 2016) Instrument: Chilean earthquake Exposure: Prenatal stress Outcome: Offspring cognitive ability (Torche, 2018)
Variation in locations and timing of policy changes or implementation	Instrument: Variation across states and time in SNAP benefit generosity, eligibility requirements, and outreach efforts Exposure: Supplemental Nutrition Assistance Program (SNAP) participation Outcome: Premature mortality (Heflin et al., 2019) Instrument: Months since pension rollout in a county Exposure: Pension enrollment and pension income Outcome: Depressive symptoms (Chen et al., 2019)
Variation in timing and locations of disasters	Instrument: Famine Exposure: Hunger early in life Outcome: Later-life health (van den Berg et al., 2016)
Timing of delivery of program benefits	Potential instrument: time of month (first two weeks vs. second two weeks) Exposure: SNAP disbursement (financial assistance for food purchases) Outcome: Types of food purchased (Franckle et al., 2019)
Eligibility cutoffs for resources based on age, income level, or other factors	Instrument: Age in years (under 18, 18 to 20, 21 and older) Exposure: Access to legal handgun purchase (varies by state policy) Outcome: Suicide (Raifman et al., 2020)
Residential moves	Instrument: Location change due to military transfer Exposure: Air pollutants Outcome: Children's health (Lleras-Muney, 2010)
Peer groups	Instrument: Assigned police officer peer group Exposure: Misconduct of peer officers Outcome: Officer misconduct (Quispe-Torreblanca & Stewart, 2019)
Genetic variants (Mendelian Randomization)	Instrument: Genetic variants associated with trauma Exposure: Trauma exposure Outcome: Psychiatric disorders (Gu et al., 2021) Instrument: Genetic variants associated with PTSD Exposure: PTSD Outcome: Alcohol use disorder (Bountress et al., 2021)
Assignment of providers or officials with different preferences	Instrument: Assigned clinicians with different treatment preferences Exposure: Nonsteroidal anti-inflammatory drugs (Brookhart & Schneeweiss, 2007); Antipsychotic medication (Rassen et al., 2009) Outcomes: Gastrointestinal toxicity (Brookhart & Schneeweiss, 2007); death (Rassen et al., 2009)

(*table continues*)

Table 2 (*continued*)

Instrument type	Examples
	Instrument: Assigned judges with different propensities for leniency Exposure: Incarceration length (Kling, 2006); prosecution and conviction for a crime (Gifford et al., 2017) Outcomes: Labor market participation (Kling, 2006); recidivism and child maltreatment (Gifford et al., 2017)
Symptom cutoffs for receiving a treatment	Instrument: Post-Traumatic Stress Disorder severity score Exposure: Trauma-specific cognitive behavioral therapy (CBT) vs. brief CBT skills intervention Outcome: Change in PTSD symptoms (CATS Consortium, 2010)
Capacity limits for patients in health care settings	Instrument: Objectively-measured hospital ward overcrowding Exposure: Job demands among nurses Outcome: Absence from work with a psychiatric diagnosis (Kivimäki et al., 2010)

conditions on relevant outcomes. Studies have also used arbitrary assignment of clinicians who may have different preferences for certain treatment modalities to understand the effects of those different treatments (Brookhart & Schneeweiss, 2007; Rassen et al., 2009), as well as arbitrary assignment of judges with different propensities for leniency, to understand the consequences of different sentencing decisions (Gifford et al., 2017; Kling, 2006).

IV analyses are often applied in the context of regression discontinuity (RD). RD designs are relevant when there is a sudden, large discontinuity in the likelihood of exposure above or below a threshold in a "forcing" variable. Examples of forcing variables include age, when the availability of a resource begins at a specific age (e.g., eligibility to purchase a handgun [Raifman et al., 2020]); symptom severity scores, when such scores are used to determine treatment eligibility (CATS Consortium, 2010); household income, when there is a sharp eligibility cutoff at a certain income level; and dates of events, when such events generate new exposures (Pesko & Baum, 2016; Torche, 2018). RD designs take advantage of the principle that individuals immediately above and below the discontinuity threshold would likely otherwise have very similar outcomes but have very different chances of exposure. Thus, they offer ideal comparison groups to estimate the effect of exposure. Some discontinuities are considered "sharp" in that the forcing variable perfectly determines exposure. With sharp discontinuities, outcomes for the groups immediately above and below the cutoff can be compared directly (without IVs) to estimate the effect of exposure. In contrast, many discontinuities are imperfect or what the literature calls "fuzzy." For example, some individuals with symptom scores below the threshold for treatment nonetheless receive treatment and many with symptom scores above the threshold still do not receive treatment. Fuzzy discontinuities can be analyzed using IV methods: the discontinuity in the forcing variable is the instrument (Oldenburg et al., 2016).

Estimating and Interpreting Causal Effects With an Instrumental Variable

Estimation

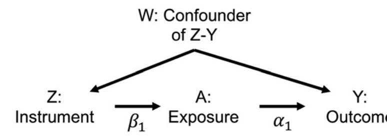
When the three conditions for a valid instrument are met, the investigator can leverage the statistical associations between the instrument, the exposure, and the outcome to estimate the causal effect of the exposure on the outcome. To offer intuition, here we describe estimation concepts for the simplest setting (continuous normally distributed variables, linear effects, linear models). Extensions are addressed in the last paragraph of this section and the additional resources section.

The instrumental variable is used to extract the variation in the exposure that is independent of the confounders. Since the instrument (Z) only affects the outcome (Y) through the exposure (A), the effect of the instrument on the outcome (θ_{ZY}) reflects both the effect of the instrument on the exposure (θ_{ZA}), and the effect of the exposure on the outcome (θ_{AY}) (i.e., $\theta_{ZY} = \theta_{ZA} * \theta_{AY}$). The causal effect of interest is θ_{AY}, so this equation can be rearranged as $\theta_{AY} = \frac{\theta_{ZY}}{\theta_{ZA}}$. Thus, to estimate the unconfounded causal effect of the exposure on the outcome, we must simply estimate the effect of the instrument on the outcome and the effect of the instrument on the exposure. In our draft lottery example, assuming the draft lottery is a valid IV for evaluating the effect of military service on mortality, if we can estimate the effect of the draft on mortality, and the effect of the draft on military service, then we can estimate the causal effect of military service on mortality.

If no other covariates need to be controlled to fulfill the required conditions, then estimating θ_{ZY} and θ_{ZA} can be as simple as applying two ordinary least squares regressions: one of the outcome on the instrument, and one of the exposure on the instrument. The ratio of these two quantities is the Wald estimator of the causal effect $\theta_{AY} = \frac{\theta_{ZY}}{\theta_{ZA}}$. In practice, however, investigators usually apply two-stage least squares (2SLS) regression to estimate valid standard errors, increase statistical power, and adjust for confounders of the instrument–outcome relationship. In 2SLS (Box 1), the exposure is first modeled as a function of the instrument and any necessary covariates. An F test of the instrument in this first-stage model is often used to verify that the relevance assumption is met (typically, $F > 10$). The fitted regression is then used to generate predicted values of the exposure for each individual in the sample. In this way, the variation in the exposure A is partitioned into (a) the variation that is related to the instrument and measured covariates (captured by the predicted exposure values, $\hat{A}$) and (b) the variation in the exposure that is related to unmeasured factors, including exposure–outcome confounders (captured by the residuals of the first-stage regression, ϵ). In 2SLS models, the latter (and hence, exposure–outcome confounding) is discarded, and in the second stage the outcome is modeled as a function of the predicted exposure values and necessary covariates. The coefficient on $\hat{A}$ in the second-stage regression (α_1) can be interpreted as the causal effect of the exposure on the outcome. Packaged functions (available in standard statistical software such as Stata, R, SAS, and SPSS) should be used to implement 2SLS because robust standard errors are needed to account for the estimation of A in the

Box 1
Two-Stage Least Squares Regression

Stage 1 regression equation: $A = \beta_0 + \beta_1 Z + \beta_2 W + \epsilon$
Stage 2 regression equation: $Y = \alpha_0 + \alpha_1 \hat{A} + \alpha_2 W + \partial$



Variable or parameter	Meaning	Simplified example: Effect of natural disaster–related housing damage on cognitive function (Hikichi et al., 2016)
Z	Instrumental variable	Distance to coast
A	Exposure	Housing damage
Y	Outcome	Cognitive function
W	Confounders of the instrument–outcome relationship	Confounders of the relationship between distance-to-coast and cognitive function (e.g., wealth)
$\hat{A}$	Predicted exposure values from the first-stage regression	Predicted levels of housing damage from first-stage regression
ϵ	First-stage regression residual (captures variation in the exposure unrelated to Z or W)	Variation in housing damage that is unrelated to distance-to-coast or confounders
α_1	Local average treatment effect (LATE) estimate of the causal effect of A on Y	LATE estimate of the causal effect of housing damage on cognitive function

first stage. Resources with sample code for software implementation are provided in the section "Sample code."

Interpreting Estimates From Instrumental Variable Analyses

How does one interpret coefficients from IV analyses (θ_{AY} or α_1)? If the effect of the exposure on the outcome is identical for everyone in the study population, then the IV estimate can be interpreted as the population average treatment effect (PATE)—for example, the difference in the mortality rate if everyone in the study population served in the military versus if no one in the study population served in the military. Confounder–control studies commonly estimate the PATE. However, homogeneous causal effects are rarely plausible. For example, the magnitude of the effect of military service on mortality likely depends on factors such as age, social resources and supports, and physical and mental health status prior to service (Vable et al., 2016). Thus, a fourth assumption is typically required for interpretation. There are several options for this fourth assumption, and the researcher can choose the assumption that seems most plausible in their setting. The option introduced in foundational work and which continues to be most popular is referred to as *monotonicity*: that the IV does not have an opposite direction of effect for any two individuals in the population (Swanson & Hernán, 2018). Monotonicity is required to estimate the LATE. In the context of the draft lottery example, if being assigned an early draft number *increased* the likelihood that some men served in the military, we must assume that there are no men for whom being assigned an early draft number *decreased* the likelihood of military service. It does not violate the monotonicity assumption if some men are simply unaffected by their draft lottery position. Like exclusion and exchangeability, monotonicity cannot be tested and its plausibility has to be judged substantively. Alternatives to monotonicity have been developed and may be more appealing for some situations (Baiocchi et al., 2014; Glymour & Swanson, 2021; Hernán & Robins, 2006; Small et al., 2017). For example, one can assume that the effect of the treatment on the outcome does not differ among treated individuals with different values of the IV; such an assumption would then estimate the effect of the treatment on the treated (rather than the LATE; Hernán & Robins, 2006). Under the relevance, exclusion, and exchangeability conditions, any association between the IV and the outcome implies a causal effect of the exposure on the outcome (i.e., tests the sharp null hypothesis that the exposure has no effect on the outcome for anyone in the population). Without the monotonicity assumption, the quantity estimated by an IV analysis cannot be interpreted as an average causal effect (Angrist et al., 1996), although one might be able to put bounds on the LATE or PATE (Hernán & Robins, 2020).

Assuming monotonicity, the estimated causal effect θ_{AY} or α_1 can be interpreted as the local average treatment effect (LATE)—that is, the effect of the variation in the exposure that is induced by the instrument. Said another way, IV analyses estimate the effect of the exposure on the outcome specifically among those individuals whose exposure is actually changed by the instrument (although who is actually affected is unknown). During the Vietnam War, some men volunteered for military service regardless of their place in the draft lottery; other men would not have served regardless of their draft lottery position (e.g., conscientious objectors). With monotonicity, an IV estimate using the draft lottery as an instrument to estimate the effect of military service on mortality is interpreted as the effect of service for men whose military service was determined by their position in the draft lottery.

Extensions

Statistical methods for IV estimation are an active area of research, and progress is extending the settings to which IVs are applicable and delivering tools to evaluate the validity of IVs (see additional resources section). This work promises to make IV methods more useful. For this introduction, several additional points are noteworthy. First, the exact estimation strategy required depends on the coding of the instrument, treatment, and outcome. If causal effects are nonlinear, alternative analytic tools are needed. Second, IVs can also be implemented using different data sets for each regression stage (i.e., two-sample IVs; Angrist & Krueger, 1992). This approach is especially useful when the exposure and outcome are not available in the same dataset or when seeking to enhance statistical power by using one larger dataset. Two-sample IVs require the assumption that the effect of the instrument on exposure is the same in the two samples and entail special standard error calculations (Inoue & Solon, 2010). If the effect of the exposure on the outcome diverges between the two samples, the IV estimate corresponds with the effect in the *outcome* sample. Finally, variations of IVs can be implemented for study designs that involve matching, time-to-event or survival outcomes, case–control sampling, multiple instruments for a single exposure variable, and to estimate parameters other than the LATE.

Cautions for Instrumental Variable Analyses

A primary challenge in using IVs to estimate causal effects in observational settings is identifying a valid instrument (i.e., a variable that meets all the required conditions). Like for confounder–control when we never truly know if we have correctly identified, measured, and controlled all confounders, for instrument-based methods, we never truly know all of the factors that determine exposure or whether all of the IV conditions are met. Many factors that appear at first to be promising instruments fail to meet one or more conditions on closer examination. For example, seemingly arbitrary variation in where and when restrictive immigration policies were adopted across states may seem to be a promising instrument to study the impacts of immigration rates on migration-related trauma. However, restrictive immigration policies might influence migration-related trauma not only by restricting immigration but also by creating a hostile social environment—a violation of the exclusion condition. Immigration policies would not be valid instruments unless one could measure and account for the hostile social environment in the analysis.

Weak instruments—when the instrument-exposure association is “small” or the F statistic for this association is “small” (Hernán & Robins, 2020)—can pose important problems for IV estimation. Weak instruments can produce severely biased effect estimates and lead to uninformatively wide or erroneously small confidence intervals (Andrews et al., 2019; Hernán & Robins, 2020). Weak instruments amplify bias owing to violations of exclusion or exchangeability, because a small and imprecise estimate of the instrument-exposure relationship (the denominator in the causal effect estimate of $\frac{\theta_{ZY}}{\theta_{ZA}}$) magnifies any bias in the instrument–outcome relationship (the numerator; Hernán & Robins, 2020). Use of multiple instruments and small sample sizes exacerbate these issues and can further amplify bias (Angrist & Krueger, 2001; Bound et al., 1995; Crown et al., 2011; Hernán & Robins, 2020). Conducting IV analyses only for very strong instruments (Lee et al., 2021) or applying weak IV-robust estimation (Andrews et al., 2019) may help remedy some of these issues.

Even with IVs that have strong associations with the exposure, small violations of the exclusion, exchangeability, and monotonicity conditions can introduce unintuitively large biases in unpredictable directions. Regardless of sample size, small departures from ideal conditions can result in IV estimates that are more biased than even a naïve estimate unadjusted for any confounders (Angrist & Krueger, 2001; Crown et al., 2011; Ding et al., 2017; Hernán & Robins, 2020; Martens et al., 2006). Exclusion is commonly violated if a continuous or multicategory exposure is coarsened to a dichotomous variable (Hernán & Robins, 2020). Exchangeability can be violated not only by confounders of the instrument–outcome relationship but also by selection bias—for example if individuals with certain levels of the exposure are excluded (Swanson et al., 2015). Monotonicity is not always a reasonable assumption in observational settings (Hernán & Robins, 2020), but estimates can be substantially biased when a small portion of the study population violates monotonicity (Swanson & Hernán, 2018). For these reasons, testing for weak instruments (Bound et al., 1995; Stock & Yogo, 2002) and conducting falsification tests and sensitivity analyses (Labrecque & Swanson, 2018; Pizer, 2016) are essential and can provide practical assessment of the impact of violating IV conditions. Applications of genetic IVs (Mendelian Randomization studies) have prompted a flourishing of methodologic developments for IV analyses, including methods that are valid under weaker assumptions (Sanderson et al., 2022). Some of these developments are also relevant for IVs arising from policy or other nongenetic sources. Additional limitations of IV methods are noted throughout the next section in contrast with confounder–control approaches.

Strengths and Limitations of Instrumental Variable Approaches Compared With Confounder–Control

We contrast the strengths and limitations of IV and confounder–control approaches with respect to the four types of validity (Matthay et al., 2019; Shadish et al., 2002).

Internal validity means that the estimated association in the study data matches the true causal effect of the exposure on the outcome for the people in the study. Between IVs and confounder–control, the approach that is more internally valid depends on which set of assumptions is more plausible for a given study context. Confounder–control is useful when adjusting for all confounders seems feasible, when there is an important research question for which no valid instrument can be identified, or when a researcher can make improvements over previous studies in identifying, measuring, or controlling confounders. However, for many stress and trauma applications, the confounding pathways are complex and unfold over long time horizons. For example, internal validity is often threatened by reverse causation, wherein the designated exposure and outcome mutually affect one another over time. Factors with bidirectional relationships include psychopathology, substance use, traumatic life experiences, and physical health status. Because instruments are theoretically external to such evolving processes, IV approaches may be less vulnerable to this bias. Yet for the many cases when no valid instrument can be identified, confounder–control approaches offer a way forward, and the potential magnitude of unmeasured confounding can be bounded with quantitative bias analysis tools (Lash et al., 2014; VanderWeele & Ding, 2017).

IVs are appealing when measuring all confounders seems infeasible but a valid instrument exists, either as-is or after adjusting for measured variables. However, for many important research questions, no valid instrument exists. Moreover, bias resulting from failure to meet the required conditions often appears to be greater for IVs than for confounder–control (see previous section). For these reasons, triangulating evidence from both confounder–control and IV approaches is likely beneficial.

Statistical conclusion validity involves using appropriate techniques for making statistical inferences about the relations between variables. To achieve statistical conclusion validity, researchers must rule out random error, meet the assumptions of the selected statistical model, and account for uncertainty in both stages of the IV estimation procedure. Because IV estimates depend on the subset of the study population whose exposure was actually changed by the instrument (even if which specific people are affected is not known), low statistical power and imprecise estimates often challenge IV approaches by increasing the likelihood that results are due to random error. For example, hospital ward overcrowding has been used as an instrument to study the causal effect of nurses' job demands on absence from work with a psychiatric diagnosis (Kivimäki et al., 2010). Because overcrowding affected job demands only for some nurses, the study sample size was effectively limited to those nurses, leading to imprecise causal effect estimates in the IV analysis. If the researchers had instead applied a confounder–control approach to the same dataset, the estimates would have been more precise. In IV analyses, precision can be enhanced using two-sample IVs (Angrist & Krueger, 1992), matching techniques (Baiocchi et al., 2012), or samples with greater proportions of individuals affected by the instrument—for example, using social media-based recruitment strategies (Schneider & Harknett, 2022).

Construct validity relates to whether the researcher has measured what they intended to measure and is a central challenge in much stress and trauma research. Statistical validity comes into tension with construct validity. For example, studies based on large administrative data sets often have more statistical power but less detailed measurements, whereas smaller studies have less statistical power but can afford more and higher quality measurements. Because instrument-based studies intrinsically have lower power, they more often rely on administrative data sets and may have less valid measures than confounder–control studies. For example, large data sets grounded in electronic health records offer promise for instrument-based stress and trauma research (Gradus et al., 2022; Weissman et al., 2020). Such records can include detailed psychopathology diagnoses, but they typically lack high-quality survey-based measures or screeners used to accurately identify symptoms and diagnoses among people not seen by mental health providers. Thus, measures of psychopathology may be highly specific but not sensitive. Approaches to addressing this challenge include taking detailed measurements on subsamples (Svensson et al., 2015) and big data initiatives that link detailed measurements with administrative records such as the U.K. Biobank and U.S.'s All of Us.

Two-sample IV methods can also help when high-quality measures of the exposure are available for only a small number of people. For example, the association between an external event and stress might be estimated in one relatively small data set with excellent stress measures. Then, the association between the external event and outcomes of interest could be estimated in another, ideally much larger data set. Information from the two data sources can be combined to derive an IV estimate of the effect of stress on the outcome.

External validity refers to the extent to which results can be generalized to people outside the study, different versions of the exposure, and other settings. Using an IV approach has important implications for generalizability because with the usual assumptions, the IV estimate only refers to the (unknown) subset of study participants whose exposure was changed by the instrument. This is a major limitation, because researchers nearly always aim to produce results that apply to people beyond the study. Confounder–control studies, which are commonly conducted in diverse or population-representative samples, are often better-positioned for generalization. Confounder–control studies can also readily examine heterogeneity in causal effects across study subgroups. In contrast, assessing heterogeneity in causal effects from IV studies requires that the instrument influences exposure to at least some degree for all subgroups; prior research has shown that this is not always the case (Lleras-Muney, 2002).

Despite its typical limited generalizability, IV approaches can still be informative, particularly when considering the delivery of services or treatments to people in need. For example, treatments or interventions for PTSD are unlikely to be delivered to people with no symptoms; of greater interest is whether people just above versus just below the symptom score cutoff for treatment eligibility have different outcomes (CATS Consortium, 2010). This is the estimate that an IV approach delivers. Additionally, alternative IV assumptions and estimation methods can deliver parameters relevant to broader populations (Glymour & Swanson, 2021).

Conclusions

Accurately estimating causal effects in observational studies is critical to strengthening evidence on the causes and consequences of stress and trauma outcomes and to informing interventions. If the measured association between an exposure and an outcome is biased away from the true causal effect because of unmeasured confounders, then interventions premised on this research will not yield the desired results. For example, if the association between PTSD and ischemic heart disease is not causal but rather reflects that people with prior combat exposure are more likely to have both PTSD and heart disease (Ebrahimi et al., 2021), then expanding access to PTSD treatment will not reduce rates of heart disease, although the research may have seemed to indicate otherwise.

Researchers must often rely on observational studies designed to minimize confounding bias. To date, most observational stress and trauma research has relied on confounder–control to address confounding, but IV approaches offer a promising alternative that could complement existing research approaches. No single study can provide the optimal degree of internal and external validity, power, and measurement, or provide inferences that are relevant to all populations and subgroups. Thus, using diverse study designs relying on alternative assumptions helps to ensure that different studies cover each other's weaknesses. Enhanced interdisciplinary collaboration between stress and trauma researchers with strong substantive expertise and econometricians with strong training in IV methods may facilitate the identification of new, valid, and useful instruments. Incorporating IV methods into the stress and trauma toolkit could therefore provide opportunities to enhance the rigor of causal research in the field.

Additional Resources for Learning More About Instrumental Variable Methods

- Angrist, JD, Pischke JS. *Mostly harmless econometrics*. Princeton University Press; 2008 Dec 15.
- Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. (1996). *Journal of the American Statistical Association*, *91*(434), 444–455.
- Angrist J. D., & Krueger A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15*(4), 69–85.
- Baiocchi M, Cheng J, Small DS. (2014). Instrumental variable methods for causal inference. *Statistics in Medicine, 33*(13), 2297–2340.
- Burgess S, Small DS, Thompson SG. (2017). A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research*, *26*(5), 2333–2355.
- Greenland S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, *29*(4), 722–729.
- Glymour, M. M., & Swanson, S. A. (2021). Instrumental variables and quasi-experimental approaches. In *Modern Epidemiology* (4th ed.). Wolters Kluwer.
- Glymour, M. M., Tchetgen Tchetgen E. J., & Robins J. M. (2012). Credible Mendelian randomization studies: Approaches for evaluating the instrumental variable assumptions. *American Journal of Epidemiology*, *175*(4), 332–339.
- Imbens G. W. (2014). Instrumental variables: An econometrician's perspective. *Statistical Science, 29*(30), 323–358.
- Harris K. M., & Remler D. K. (1998). Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. *Health Services Research, 33*(5 Pt 1), 1337–1360.
- Labrecque, J., & Swanson, S. A. (2018). Understanding the assumptions underlying instrumental variable analyses: A brief review of falsification strategies and related tools. *Current Epidemiology Reports*, *5*(3), 214–220.
- Pierce BL, Burgess S. (2013). Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *American Journal of Epidemiology*, *178*, 1177–1184.
- Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. (2009). Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *Journal of Clinical Epidemiology*, *62*(12), 1226–1232.
- Rassen, J. A., Brookhart, M. A., Glynn, R. J., Mittleman, M. A., & Schneeweiss, S. (2009). Instrumental variables II: Instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *Journal of Clinical Epidemiology*, *62*(12), 1233–1241.
- Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, Palmer T, Schooling CM, Wallace C, Zhao Q, Davey Smith G. (2022). Mendelian randomization. *Nature Reviews Methods Primers*, *2*(1), 1–21.
- Steiner PM, Kim Y, Hall CE, Su D. (2015). Graphical models for quasi-experimental designs. *Sociological Methods & Research*, *46*(2), 155–188.
- Swanson SA, Hernan MA. (2013). Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology*, *24*, 370–374.
- Swanson SA, Hernán MA. (2014). Think globally, act globally: An epidemiologist's perspective on instrumental variable estimation. *Statistical Science, 29*(3), 371.

Sample Code

- Baiocchi M, Cheng J, Small DS. (2014). Instrumental variable methods for causal inference. *Statistics in Medicine, 33*(13), 2297–2340.
- Baum CF, Schaffer ME, Stillman S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *The Stata Journal*, 7(4), 465–506. https://journals.sagepub.com/doi/pdf/10.1177/1536867X0800700402
- Cunningham S. (2021). Chapter 9: Difference-in-differences. In *Causal inference: The mixtape*. Yale University Press. https://mixtape.scunning.com/instrumental-variables.html
- Hernan MA, Robins JM. (2020). Chapter 16: Instrumental variables estimation. In *Causal inference: What if* (1st ed., pp.193–208). Chapman & Hall/CRC. https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
- Huntington-Klein N. (2021). Chapter 19: Instrumental variables. In *The effect: An introduction to research design and causality*. Chapman & Hall. https://theeffectbook.net/ch-InstrumentalVariables.html
- Ryan AM, Burgess Jr JF, Dimick JB. (2015). Why we should not be indifferent to specification choices for difference-in-differences. *Health Services Research*, 50(4), 1211–1235.

References

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, *11*(1), 727–753. https://doi.org/10.1146/annurev-economics-080218-025643

Angrist, J. D., & Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, *88*(3), 450–477.

Angrist, J. D., & Krueger, A. B. (1992). 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*, *87*(418), 328–336. https://doi.org/10.1080/01621459.1992.10475212

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *The Journal of Economic Perspectives*, *15*(4), 69–85. https://doi.org/10.1257/jep.15.4.69

Angrist, J. D., Chen, S. H., & Frandsen, B. R. (2010). Did Vietnam veterans get sicker in the 1990s? The complicated effects of military service on self-reported health. *Journal of Public Economics*, *94*(11–12), 824–837. https://doi.org/10.1016/j.jpubeco.2010.06.001

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. The. *Journal of Economic Perspectives*, *31*(2), 3–32. https://doi.org/10.1257/jep.31.2.3

Austin, A. E., Desrosiers, T. A., & Shanahan, M. E. (2019). Directed acyclic graphs: An under-utilized tool for child maltreatment research. *Child Abuse & Neglect*, *91*, 78–87. https://doi.org/10.1016/j.chiabu.2019.02.011

Baiocchi, M., Cheng, J., & Small, D. S. (2014). Instrumental variable methods for causal inference. *Statistics in Medicine*, *33*(13), 2297–2340. https://doi.org/10.1002/sim.6128

Baiocchi, M., Small, D. S., Yang, L., Polsky, D., & Groeneveld, P. W. (2012). Near/far matching: A study design approach to instrumental variables. *Health Services and Outcomes Research Methodology*, *12*(4), 237–253. https://doi.org/10.1007/s10742-012-0091-0

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is Weak. *Journal of the American Statistical Association*, *90*(430), 443–450. https://doi.org/10.1080/01621459.1995.10476536

Bountress, K. E., Wendt, F., Bustamante, D., Agrawal, A., Webb, B., Gillespie, N., Edenberg, H., Sheerin, C., Johnson, E., Polimanti, R., Amstadter, A., Amstadter, A., & The Psychiatric Genomics Consortium Posttraumatic Stress Disorder Working Group. (2021). Potential causal effect of posttraumatic stress disorder on alcohol use disorder and alcohol consumption in individuals of European descent: A Mendelian Randomization Study. *Alcoholism, Clinical and Experimental Research*, *45*(8), 1616–1623. https://doi.org/10.1111/acer.14649

Bronars, S. G., & Grogger, J. (1994). The Economic consequences of unwed motherhood: Using twin births as a natural experiment. *The American Economic Review*, *84*(5), 1141–1156.

Brookhart, M. A., & Schneeweiss, S. (2007). Preference-based instrumental variable methods for the estimation of treatment effects: Assessing validity and interpreting results. *The International Journal of Biostatistics*, *3*(1), 14. https://doi.org/10.2202/1557-4679.1072

CATS Consortium. (2010). Implementation of CBT for youth affected by the World Trade Center disaster: Matching need to treatment intensity and reducing trauma symptoms. *Journal of Traumatic Stress*, *23*(6), 699–707. https://doi.org/10.1002/jts.20594

Chen, X., Wang, T., & Busch, S. H. (2019). Does money relieve depression? Evidence from social pension expansions in China. *Social Science & Medicine*, *220*, 411–420. https://doi.org/10.1016/j.socscimed.2018.12.004

Conley, D., & Heerwig, J. (2012). The long-term effects of military conscription on mortality: Estimates from the Vietnam-era draft lottery. *Demography*, *49*(3), 841–855. https://doi.org/10.1007/s13524-012-0103-2

Crown, W. H., Henk, H. J., & Vanness, D. J. (2011). Some cautions on the use of instrumental variables estimators in outcomes research: How bias in instrumental variables estimators is affected by instrument strength, instrument contamination, and sample size. *Value in Health*, *14*(8), 1078–1084. https://doi.org/10.1016/j.jval.2011.06.009

Ding, P., VanderWeele, T. J., & Robins, J. M. (2017). Instrumental variables as bias amplifiers with general outcome and confounding. *Biometrika*, *104*(2), 291–302. https://doi.org/10.1093/biomet/asx009

Ebrahimi, R., Lynch, K. E., Beckham, J. C., Dennis, P. A., Viernes, B., Tseng, C.-H., Shroyer, A. L. W., & Sumner, J. A. (2021). Association of posttraumatic stress disorder and incident ischemic heart disease in women veterans. *JAMA Cardiology*, *6*(6), 642–651. https://doi.org/10.1001/jamacardio.2021.0227

Fernandez, C. A., Vicente, B., Marshall, B. D., Koenen, K. C., Arheart, K. L., Kohn, R., Saldivia, S., & Buka, S. L. (2017). Longitudinal course of disaster-related PTSD among a prospective sample of adult Chilean natural disaster survivors. *International Journal of Epidemiology*, *46*(2), 440–452. https://doi.org/10.1093/ije/dyw094

Franckle, R. L., Thorndike, A. N., Moran, A. J., Hou, T., Blue, D., Greene, J. C., Bleich, S. N., Block, J. P., Polacsek, M., & Rimm, E. B. (2019). Supermarket purchases over the supplemental nutrition assistance program benefit month: A comparison between participants and nonparticipants. *American Journal of Preventive Medicine*, *57*(6), 800–807. https://doi.org/10.1016/j.amepre.2019.07.025

Gade, D. M., & Wenger, J. B. (2011). Combat exposure and mental health: The long-term effects among U.S. Vietnam and Gulf War veterans. *Health Economics*, *20*(4), 401–416. https://doi.org/10.1002/hec.1594

Gifford, E. J., Eldred, L. M., Mccutchan, S. A., & Sloan, F. A. (2017). Prosecution, Conviction, and Deterrence in Child Maltreatment Cases. *Criminal Justice and Behavior*, *44*(10), 1262–1280. https://doi.org/10.1177/0093854817727795

Glymour, M. M., & Swanson, S. A. (2021). Instrumental variables and quasi-experimental approaches. In T. L. Lash, T. J. VanderWeele, S. Haneuse, & K. J. Rothman (Eds.), *Modern Epidemiology* (Fourth; pp. 677–710). Wolters Kluwer.

Gradus, J. L., Rosellini, A. J., Szentkúti, P., Horváth-Puhó, E., Smith, M. L., Galatzer-Levy, I., Lash, T. L., Galea, S., Schnurr, P. P., & Sørensen, H. T. (2022). Using Danish national registry data to understand psychopathology following potentially traumatic experiences. *Journal of Traumatic Stress*, *35*(2), 619–630. https://doi.org/10.1002/jts.22777

Gu, D., Ou, S., & Liu, G. (2021). Causal association of trauma with subsequent psychiatric disorder: A Mendelian randomization study. *Research Square*. https://doi.org/10.21203/rs.3.rs-870584/v1

Heflin, C. M., Ingram, S. J., & Ziliak, J. P. (2019). The effect of the supplemental nutrition assistance program on mortality. *Health Affairs (Project Hope)*, *38*(11), 1807–1815. https://doi.org/10.1377/hlthaff.2019.00405

Hernán, M. A., & Robins, J. M. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology*, *17*(4), 360–372. https://doi.org/10.1097/01.ede.0000222409.00878.37

Hernán, M. A., & Robins, J. M. (2020). Instrumental variables estimation. In *Causal inference: What if* (pp. 197–214). Chapman & Hall/CRC.

Hikichi, H., Aida, J., Kondo, K., Tsuboya, T., Matsuyama, Y., Subramanian, S. V., & Kawachi, I. (2016). Increased risk of dementia in the aftermath of the 2011 Great East Japan Earthquake and Tsunami. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(45), E6911–E6918. https://doi.org/10.1073/pnas.1607793113

Inoue, A., & Solon, G. (2010). Two-sample instrumental variables estimators. *The Review of Economics and Statistics*, *92*(3), 557–561. https://doi.org/10.1162/REST_a_00011

Jiang, T., Smith, M. L., Street, A. E., Seegulam, V. L., Sampson, L., Murray, E. J., Fox, M. P., & Gradus, J. L. (2021). A comorbid mental disorder paradox: Using causal diagrams to understand associations between posttraumatic stress disorder and suicide. *Psychological Trauma: Theory, Research, Practice, and Policy*, *13*(7), 725–729. https://doi.org/10.1037/tra0001033

Johnston, D. W., Shields, M. A., & Siminski, P. (2016). Long-term health effects of Vietnam-era military service: A quasi-experiment using Australian conscription lotteries. *Journal of Health Economics*, *45*, 12–26. https://doi.org/10.1016/j.jhealeco.2015.11.003

Kivimäki, M., Vahtera, J., Kawachi, I., Ferrie, J. E., Oksanen, T., Joensuu, M., Pentti, J., Salo, P., Elovainio, M., & Virtanen, M. (2010). Psychosocial work environment as a risk factor for absence with a psychiatric diagnosis: An instrumental-variables analysis. *American Journal of Epidemiology*, *172*(2), 167–172. https://doi.org/10.1093/aje/kwq094

Kling, J. R. (2006). Incarceration length, employment, and earnings. *The American Economic Review*, *96*(3), 863–876. https://doi.org/10.1257/aer.96.3.863

Labrecque, J., & Swanson, S. A. (2018). Understanding the assumptions underlying instrumental variable analyses: A brief review of falsification strategies and related tools. *Current Epidemiology Reports*, *5*(3), 214–220. https://doi.org/10.1007/s40471-018-0152-1

Lash, T. L., Fox, M. P., MacLehose, R. F., Maldonado, G., McCandless, L. C., & Greenland, S. (2014). Good practices for quantitative bias analysis. *International Journal of Epidemiology*, *43*(6), 1969–1985. https://doi.org/10.1093/ije/dyu149

Lee, D. S., McCrary, J., Moreira, M. J., & Porter, J. R. (2021). Valid t-ratio Inference for IV (Working Paper No. 29124). *National Bureau of Economic Research*. https://doi.org/10.3386/w29124

Lleras-Muney, A. (2002). Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939. *The Journal of Law & Economics*, *45*(2), 401–435. https://doi.org/10.1086/340393

Lleras-Muney, A. (2010). The needs of the army using compulsory relocation in the military to estimate the effect of air pollutants on children's health. *The Journal of Human Resources*, *45*(3), 549–590. https://doi.org/10.3368/jhr.45.3.549

Lyk-Jensen, S. V., Weatherall, C. D., & Jepsen, P. W. (2016). The effect of military deployment on mental health. *Economics and Human Biology*, *23*, 193–208. https://doi.org/10.1016/j.ehb.2016.09.005

Martens, E. P., Pestman, W. R., de Boer, A., Belitser, S. V., & Klungel, O. H. (2006). Instrumental variables: Application and limitations. *Epidemiology*, *17*(3), 260–267. https://doi.org/10.1097/01.ede.0000215160.88317.cb

Matthay, E. C., Hagan, E., Gottlieb, L. M., Tan, M. L., Vlahov, D., Adler, N. E., & Glymour, M. M. (2019). Alternative causal inference methods in population health research: Evaluating tradeoffs and triangulating evidence. *SSM - Population Health*, *10*, Article 100526. https://doi.org/10.1016/j.ssmph.2019.100526

Oldenburg, C. E., Moscoe, E., & Bärnighausen, T. (2016). Regression discontinuity for causal effect estimation in epidemiology. *Current Epidemiology Reports*, *3*(3), 233–241. https://doi.org/10.1007/s40471-016-0080-x

Osypuk, T. L., Schmidt, N. M., Kehm, R. D., Tchetgen Tchetgen, E. J., & Glymour, M. M. (2019). The price of admission: Does moving to a low-poverty neighborhood increase discriminatory experiences and influence mental health? *Social Psychiatry and Psychiatric Epidemiology*, *54*(2), 181–190. https://doi.org/10.1007/s00127-018-1592-0

Palmer, M., Nguyen, C. V., Mitra, S., Mont, D., & Groce, N. E. (2019). Long-lasting consequences of war on disability. *Journal of Peace Research*, *56*(6), 860–875. https://doi.org/10.1177/0022343319846545

Pearl, J. (2009). *Causality*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Pesko, M. F., & Baum, C. F. (2016). The self-medication hypothesis: Evidence from terrorism and cigarette accessibility. *Economics and Human Biology*, *22*, 94–102. https://doi.org/10.1016/j.ehb.2016.03.007

Pizer, S. D. (2016). Falsification testing of instrumental variables methods for comparative effectiveness research. *Health Services Research*, *51*(2), 790–811. https://doi.org/10.1111/1475-6773.12355

Quispe-Torreblanca, E. G., & Stewart, N. (2019). Causal peer effects in police misconduct. *Nature Human Behaviour*, *3*(8), 797–807. https://doi.org/10.1038/s41562-019-0612-8

Raifman, J., Larson, E., Barry, C. L., Siegel, M., Ulrich, M., Knopov, A., & Galea, S. (2020). State handgun purchase age minimums in the U.S. and adolescent suicide rates: Regression discontinuity and difference-in-differences analyses. *BMJ*, *370*, m2436. https://doi.org/10.1136/bmj.m2436

Rassen, J. A., Brookhart, M. A., Glynn, R. J., Mittleman, M. A., & Schneeweiss, S. (2009). Instrumental variables II: Instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *Journal of Clinical Epidemiology*, *62*(12), 1233–1241. https://doi.org/10.1016/j.jclinepi.2008.12.006

Sanderson, E., Richardson, T. G., Morris, T. T., Tilling, K., Davey Smith, G., Munafò, M. R., Palmer, T., Schooling, C. M., Wallace, C., Zhao, Q., & Davey Smith, G. (2022). Estimation of causal effects of a time-varying exposure at multiple time points through multivariable mendelian randomization. *PLoS Genetics*, *18*(7), e1010290. https://doi.org/10.1371/journal.pgen.1010290

Scherrer, J. F., Salas, J., Schneider, F. D., Friedman, M. J., van den Berk-Clark, C., Chard, K. M., Norman, S. B., Lustman, P. J., Tuerk, P., Schnurr, P. P., & Cohen, B. E. (2020). PTSD improvement and incident cardiovascular disease in more than 1000 veterans. *Journal of Psychosomatic Research*, *134*, Article 110128. https://doi.org/10.1016/j.jpsychores.2020.110128

Schneider, D., & Harknett, K. (2022). What's to like? Facebook as a tool for survey data collection. *Sociological Methods & Research*, *51*(1), 108–140. https://doi.org/10.1177/0049124119882477

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin and Company.

Singhal, S. (2019). Early life shocks and mental health: The long-term effect of war in Vietnam. *Journal of Development Economics*, *141*, 102244. https://doi.org/10.1016/j.jdeveco.2018.06.002

Small, D. S., Tan, Z., Ramsahai, R. R., Lorch, S. A., & Brookhart, M. A. (2017). Instrumental variable estimation with a stochastic monotonicity assumption. *Statistical Science*, *32*(4), 561–579. https://doi.org/10.1214/17-STS623

Stock, J. H., & Yogo, M. (2002). *Testing for weak instruments in linear IV regression* (No. 0284). NBER. https://doi.org/10.3386/t0284

Sumner, J. A., Nishimi, K. M., Koenen, K. C., Roberts, A. L., & Kubzansky, L. D. (2020). Posttraumatic Stress Disorder and Inflammation: Untangling Issues of Bidirectionality. *Biological Psychiatry*, *87*(10), 885–897. https://doi.org/10.1016/j.biopsych.2019.11.005

Svensson, E., Lash, T. L., Resick, P. A., Hansen, J. G., & Gradus, J. L. (2015). Validity of reaction to severe stress and adjustment disorder diagnoses in the Danish Psychiatric Central Research Registry. *Clinical Epidemiology*, *7*, 235–242. https://doi.org/10.2147/CLEP.S80514

Swanson, S. A., & Hernán, M. A. (2018). The challenging interpretation of instrumental variable estimates under monotonicity. *International Journal of Epidemiology*, *47*(4), 1289–1297. https://doi.org/10.1093/ije/dyx038

Swanson, S. A., Robins, J. M., Miller, M., & Hernán, M. A. (2015). Selecting on treatment: A pervasive form of bias in instrumental variable analyses. *American Journal of Epidemiology*, *181*(3), 191–197. https://doi.org/10.1093/aje/kwu284

Syed Sheriff, R., Van Hooff, M., Malhi, G., Grace, B., & McFarlane, A. (2019). Associations among childhood trauma, childhood mental disorders, and past-year posttraumatic stress disorder in military and civilian men. *Journal of Traumatic Stress*, *32*(5), 712–723. https://doi.org/10.1002/jts.22450

Torche, F. (2018). Prenatal exposure to an acute stressor and children's cognitive outcomes. *Demography*, *55*(5), 1611–1639. https://doi.org/10.1007/s13524-018-0700-9

Tsai, A. C., & Venkataramani, A. S. (2015). The causal effect of education on HIV stigma in Uganda: Evidence from a natural experiment. *Social Science & Medicine*, *142*, 37–46. https://doi.org/10.1016/j.socscimed.2015.08.009

Vable, A. M., Canning, D., Glymour, M. M., Kawachi, I., Jimenez, M. P., & Subramanian, S. V. (2016). Can social policy influence socioeconomic disparities? Korean War GI Bill eligibility and markers of depression. *Annals of Epidemiology*, *26*(2), 129–135.e3. https://doi.org/10.1016/j.annepidem.2015.12.003

van den Berg, G. J., Pinger, P. R., & Schoch, J. (2016). Instrumental variable estimation of the causal effect of hunger early in life on health later in life. *Economic Journal (London)*, *126*(591), 465–506. https://doi.org/10.1111/ecoj.12250

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: Introducing the e-value. *Annals of Internal Medicine*, *167*(4), 268–274. https://doi.org/10.7326/M16-2607

Weissman, M. M., Pathak, J., & Talati, A. (2020). Personal life events-A promising dimension for psychiatry in electronic health records. *JAMA Psychiatry*, *77*(2), 115–116. https://doi.org/10.1001/jamapsychiatry.2019.3217

White, J. S., Hamad, R., Li, X., Basu, S., Ohlsson, H., Sundquist, J., & Sundquist, K. (2016). Long-term effects of neighbourhood deprivation on diabetes risk: Quasi-experimental evidence from a refugee dispersal policy in Sweden. *The Lancet. Diabetes & Endocrinology*, *4*(6), 517–524. https://doi.org/10.1016/S2213-8587(16)30009-2

Received March 4, 2022
Revision received July 11, 2022
Accepted July 25, 2022 ■