JAMA Psychiatry | Review

Applying Causal Inference Methods in Psychiatric Epidemiology
A Review

Henrik Ohlsson, PhD; Kenneth S. Kendler, MD

IMPORTANCE Associations between putative risk factors and psychiatric and substance use disorders are widespread in the literature. Basing prevention efforts on such findings is hazardous. Applying causal inference methods, while challenging, is central to developing realistic and potentially actionable etiologic models for psychopathology.

OBSERVATIONS Causal methods can be divided into randomized clinical trials (RCTs), natural experiments, and statistical models. The first 2 approaches can potentially control for both known and unknown confounders, while statistical methods control only for known and measured confounders. The criterion standard, RCTs, can have important limitations, especially regarding generalizability. Furthermore, for ethical reasons, many critical questions in psychiatric epidemiology cannot be addressed by RCTs. We review, with examples, methods that try to meet as-if randomization assumptions, use instrumental variables, or use pre-post designs, regression discontinuity designs, or co-relative designs. Each method has strengths and limitations, especially the plausibility of as-if randomization and generalizability. Of the large family of statistical methods for causal inference, we examine propensity scoring and marginal models, which are best applied to samples with strong predictors of risk factor exposure.

CONCLUSIONS AND RELEVANCE Causal inference is important because it informs etiologic models and prevention efforts. The view that causation can be definitively resolved only with RCTs and that no other method can provide potentially useful inferences is simplistic. Rather, each method has varying strengths and limitations. We need to avoid the extremes of overzealous causal claims and the cynical view that potential causal information is unattainable when RCTs are infeasible. Triangulation, which applies different methods for elucidating causal inferences to address to the same question, may increase confidence in the resulting causal claims.

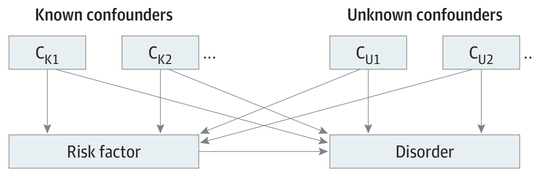
JAMA Psychiatry. doi:10.1001/jamapsychiatry.2019.3758
Published online December 11, 2019.

Author Affiliations: Center for Primary Health Care Research, Lund University, Malmö, Sweden (Ohlsson); Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond (Kendler); Department of Psychiatry, Virginia Commonwealth University, Richmond (Kendler).

Corresponding Author: Kenneth S. Kendler, MD, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, PO Box 980126, Richmond, VA 23298 (kenneth.kendler@vcuhealth.org).

Causation, long a subject of critical debate in the philosophy,[1] epidemiology,[2-4] and statistics,[5-8] is also a subject of immense practical importance. The observation of an association between a putative risk factor and a disorder provides a descriptive account of the world but gives no insights into the origins of that association or ways it might be altered. Here, we provide an overview of approaches to causal inference in psychiatric epidemiology. We seek to convey the logic of the various methods for examining average causal effects (the mean difference between individuals exposed and unexposed to an intervention in some well-defined population) and to discuss their strengths and limitations but not their detailed statistical foundation or specific suggestions of when each method should be used. Causal inference always requires methodologic assumptions[9] and rarely produces unequivocal results. Furthermore, while randomized clinical trials (RCTs) can theoretically provide findings with internal validity regarding the effect of an intervention, no single study can provide a definitive answer to the questions of causality. The fundamental problem in causal inference is distinguishing whether an observed association between a putative risk factor and a disorder, where the risk factor precedes the disorder, results from causal influences of the risk factor on the disease or arises from the influence of known and unknown confounding variables that may affect both of them (**Figure 1**). For these purposes, confounders can be usefully divisible into 2 groups: (1) unknown and/or unmeasured and (2) both known and measured.

Figure 1. A Situation in Which the Association Between the Risk Factor and the Disorder Are Confounded by Both Measured Confounders (C_{K1} and C_{K2}) and Unmeasured Confounders (C_{U1} and C_{U2})



A standard multiple regression can only control for C_{K1} and C_{K2} while a natural experiment, the co-relative design, and a randomized clinical trial also control for C_{U1} and C_{U2}.

© 2019 American Medical Association. All rights reserved.

Practical efforts to infer causality can be compared with an impossible thought experiment in which we replicate our world. In 1 replicate, research participants are exposed to a risk factor or potential cause, and in the other, they would not be exposed. We then compare the rates of disorder in the 2 worlds. This experiment is strong because all confounders (both unknown/unmeasured and known/measured) are controlled. The experimental and control study participants are the same people, and only the exposure is different. None of the research methods we are reviewing achieve this lofty goal, but each method both attempts to do so and fails in interesting ways.

Randomized Clinical Trials

The best possible equivalent to our ideal experiment is the randomized clinical trial (RCT). While RCTs typically evaluate treatments, we focus here on their application to risk factor exposures. Randomized clinical trials meet 3 criteria:

1. The response of experimental participants assigned to exposure is compared with the response of participants assigned to a nonexposed control group.
2. The assignment of participants to exposure and control groups is random.
3. The manipulation of the exposure is controlled by the researcher.

By randomizing individuals to 2 groups, RCTs attempt to divide the known and unknown confounding factors evenly across the 2 groups, so that the study groups differ systematically (at least in theory) only by risk factor exposure.[10] As in our ideal scenario, in strong RCTs, no statistical controls are needed. The causal effect of exposure equals the difference in rates of illness in the exposed and unexposed groups.

Randomized clinical trials are widely perceived as the criterion standard in causal inference and more reliable and credible than any other method. There is strength to this claim: RCTs are able to reliably account for confounding from both known and unknown sources and provide the best approach to assess internal validity. However, the special status of RCTs can be overestimated.[11-15] For example, individuals participating in RCTs are often not representative of the population typically exposed to the intervention. Therefore, establishing causality in an RCT does not guarantee extrapolation to the general population. The duration of exposures to the intervention in RCTs is often briefer than typical in the population. Even in a well-conducted RCT, attrition, nonadherence, unintentional unblinding and other postrandomization confounding, and selection biases are not uncommon. Randomized clinical trials can also be very expensive and time-consuming. Most importantly, for many critical risk factors in psychiatry for which potential causal relationships with disorders are unclear, conducting an RCT is unethical and/or impractical.

In such situations, any attempt to infer a potential causal relationship between risk factors and disease must turn to either natural experiments or statistical models.[16] These approaches have a critical difference. Natural experiments (an observational study in which the experimental variables of interest are influenced by factors outside of the researchers' control), with varying degree of coverage and confidence, can, when properly conducted and analyzed, control for many or most confounders, including those that are not known or known and not measured. Statistical models can only control for confounders that are both known and measured. There are a wide variety of kinds of natural experiments and herein we review a selection of these.

Natural Experiments

With Randomization or *As-If* Randomization

Some natural experiments closely approximate RCTs, lacking only criterion 3 (exposure controlled by the researcher). The risk-factor exposure is randomized by social or political processes not researchers. **Table 1** provides details of an example of a natural experiment using the *British Household Panel Survey*[17] that showed improvements in mental health after winning a lottery prize.

In many studies, an *as-if* rather than formal randomization process is used. The strength of the causal inference in such experiments is closely related to the degree of confidence that can be placed in the as-if random process. Dunning[4(p235-254)] has a helpful discussion of this question and focuses on the problem of self-selection into treatment groups. He recommends that researchers evaluate whether the study participants had the necessary information, incentives, and capacities to control their own assignment. He suggests the helpful concept of a "continuum of plausibility"[4] for such studies, defined by the extent to which treatment assignment is plausibly as-if random.

A classic example of such a study was done by John Snow of the 1853/1854 London, England, cholera epidemic[18] (Table 1). Major areas of London were served by 2 water companies that had typically been selected by landlords years before the epidemic, when both companies took their water downstream of the main London sewers. One year before the outbreak, one company moved their intake upstream. Death rates from cholera were 8.5 times greater in houses served by the company taking their water downstream vs upstream of the sewage discharge. Snow writes that the distribution of these 2 water services divided the London population into 2 groups "without their choice and in most cases without their knowledge."[18] In aggregate, the data presented by Snow support the plausibility of the as-if random assignment of exposure to relatively clean vs sewage-contaminated water.

Instrumental Variable Analyses

Sometimes, natural experiments with as-if random exposure turn their focus from the risk factor itself (eg, winning a lottery or consuming contaminated water) to an instrument that predicts risk factor exposure. The logic of this *instrumental variable design* is seen in **Figure 2**A.[25] Critically, the instrument affects the outcome only through its influence on the exposure. The specialized topic of mendelian randomization,[26] reviewed in *JAMA* in 2017,[26] is a particular form of instrumental variable analysis.

As detailed in Table 1, Wang et al[19] studied the association between different types of antipsychotic medications in elderly patients in Pennsylvania and risk of death in a retrospective cohort study using instrumental variable analyses. The instrument was the prescribing physician's preference for conventional or atypical antipsychotic medications (as indicated by the physicians' most recent new antipsychotic prescriptions). Their instrumental-variable

© 2019 American Medical Association. All rights reserved.

Table 1. Examples of Studies Using a Range of Methods of Causal Inference

Source	Type of Natural Experiment	Aim	Study Description and Summary of Results
Apouey and Clark[17]	Randomization through a lottery; winners of money through a lottery are randomly selected.	To measure the association between lottery income and different health measures.	The British Household Panel Survey included information about lottery winnings as well as a number of measures of general health status and mental health. The authors showed that receipt of lottery winnings had no significantly associated effect on self-assessed physical health, but was significantly associated with mental health.
Snow[18]	As-if randomization. Change of intake of drinking water for a portion of the population in London.	To investigate the sources of cholera outbreak in London, England, in 1854.	Before the cholera epidemic, large areas of London were served by 2 major water companies that both took their water from the Thames downstream of the main London sewers. The choice of householders of which company to use was typically made years before the cholera outbreak often by distant landlords. One year before the outbreak, 1 company moved their water intake upstream. Snow writes that "there is no difference either in the condition or occupation of the persons receiving the water of the different Companies."[18] The death rates from cholera were 8.5 times greater in houses served by the company taking their water downstream vs upstream of the sewage discharge.
Wang et al[19]	Instrumental variable. The instrument was the prescribing physician's preference for conventional or atypical antipsychotic medications (as indicated by his or her most recent new prescription for an antipsychotic agent).	To compare the risk of death within 180 d, <40 d, 40-79 d, and 80-180 d after the initiation of therapy with a conventional or atypical antipsychotic medication.	The authors conducted a retrospective cohort study involving 22 890 patients 65 y or older who had drug insurance benefits in Pennsylvania and who began receiving conventional or atypical antipsychotic medication between 1994 and 2003. Conventional antipsychotic medications were associated with a significantly higher risk of death than were atypical antipsychotic medications at all intervals studied.
Kendler et al[20]	Instrumental variable and co-relative design. Month of birth (the instrument) was highly associated with academic achievement but not drug abuse (disorder).	To determine whether the association between poor AA and risk of DA is influenced by potential causal processes.	Lower AA was associated with subsequent DA registration (HR per SD, 2.33; 95% CI, 2.30-2.35). Instrumental variable analysis produced an attenuated association (HR, 2.04; 95% CI, 1.75-2.33). In the co-relative design, the AA-DA association in monozygotic twins was estimated to equal 1.79 (95% CI, 1.64-1.92). Two different approaches both produced results consistent with the hypothesis that the association observed between AA and risk of DA into middle adulthood may be causal.
Kreitman[21]	Pre-post design. In 1963, the gas companies in Great Britain started decreasing the CO content of the gas.	An analysis of the declining rates of suicide between 1960 and 1971 for England and Wales and for Scotland.	Death from carbon monoxide poisoning from coal gas delivered to homes for cooking and heating was the most common form of suicide in the United Kingdom in the 1950s, with stable rates for nearly a decade. After the companies started decreasing the CO content of the gas, a sharp and nearly contemporary decline in the suicide rates were observed. When divided by form of death, the decline was entirely in those from CO poisoning. In men, no parallel increase was seen in suicides by other means, providing evidence against the substitution theory that prevention-specific methods for suicide would not reduce total suicides because distressed individuals would merely select another method of dying.

(continued)

analyses suggested that conventional antipsychotic medications were potentially causally associated with a significantly higher risk of death.

In 2018, we examined the potential causal association between academic achievement at age 16 years and risk for drug abuse[20] (Table 1) using instrumental variable analyses. Our instrument was month of birth because, with rare exceptions, all members of a school class in Sweden are born in the same year. Within such classes, academic achievement was consistently higher in the older than in the younger children. Furthermore, month of birth was not associated with drug abuse risk when controlling for academic achievement. The plausibility of the as-if randomization in this study supported the conclusion that low academic achievement was potentially causally associated with drug abuse risk.

Pre-Post Designs

Pre-post design natural experiments, sometimes called regression discontinuity designs with longitudinal specification,[27] do not usually con-

© 2019 American Medical Association. All rights reserved.

Table 1. Examples of Studies Using a Range of Methods of Causal Inference (continued)

Source	Type of Natural Experiment	Aim	Study Description and Summary of Results
Costello et al[22]	Pre-post design. Using an intervention, a casino opening that affected parts of the population.	To assess whether the high prevalence of mental illness among poor people is related with potential social causation or a social selection.	A representative population sample of 1420 rural children aged 9 to 13 y at intake were given annual psychiatric assessments for 8 y. Families were categorized into 3 groups: persistently poor, who remained poor all the way before and after the intervention; ex-poor, who were poor before the intervention and came out of poverty after the intervention; and never poor. After the intervention, the likelihood of having mental disorders among the children of ex-poor and never poor families was almost similar. The mean psychometric symptoms scores decreased significantly among the children of ex-poor families when they came out of poverty. The results of this study supported the social causation theory for externalizing symptoms in children.
Humphreys et al[23]	Pre-post design. Using an intervention, an implementation of a new law in one state but not in comparable states.	To estimate the relationship between Florida's Stand Your Ground law rates of homicide and homicide by firearm.	In 2005, Florida amended its self-defense laws to provide legal immunity to individuals using lethal force in self-defense. Prior to the Stand Your Ground law, the mean monthly homicide rate in Florida was 0.49 deaths per 100 000 and the rate of homicide by firearm was 0.29 deaths per 100 000. After accounting for underlying trends, there was an abrupt and sustained increase in the monthly homicide rate of 24.4% and in the rate of homicide by firearm of 31.6% after the law was implemented. No evidence of change was found in the analyses of comparison states.
Taylor et al[24]	Propensity score matching. Nicotine dependency together with sex, mental health, intention to quit, and several other variables were included in the propensity score calculation.	To estimate the potential causal relationship between of smoking cessation and mental health.	Using 937 individuals who had smoked for at least 3 y, the authors matched individuals on their propensity to stop smoking. Using this technique, they achieved a good match between smokers that continued smoking and those that stopped. The regression coefficient from the propensity analysis for the difference between smokers and quitters was lower than that achieved by ordinary regression methods alone, and no longer showed a significant positive association between cessation on mental health status.

Abbreviations: AA, academic achievement; CO, carbon monoxide; DA, drug abuse.

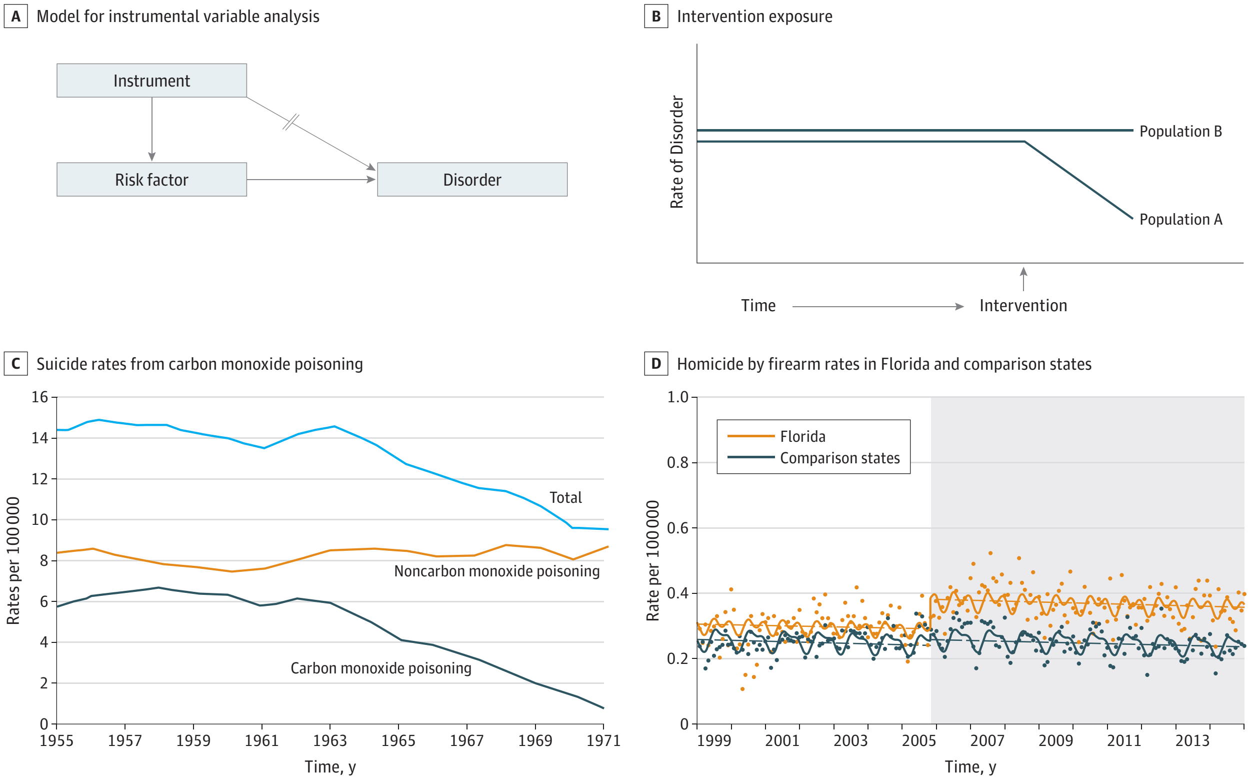
tain as-if randomization and so might be more susceptible to confounder bias. Typically, the intervention is a change in public policy, criminal law, or system of income disbursal. In the simplest pre-post design, rates of the disorder are examined in the same population before and after intervention (Figure 2B). The plausibility of causal inference depends on the stability of the historical trends and the likelihood that other confounding factors changed around the time of the intervention. A classic example of this approach is the analysis of coal-gas and suicides in the England and Wales (Table 1),[21] which associated declining suicides from carbon monoxide (CO) poisoning in coal gas with reductions in the CO content of that gas (Figure 2C). Importantly, in men, no parallel increase was seen in suicides by other means, providing evidence against the substitution theory that prevention of specific methods for suicide would not reduce suicides because distressed individuals would select another method.[28] A substantial literature that uses this method provides evidence for a causal association between alterations in tax rates on alcohol with the availability or pricing of alcohol beverages and the closing hours of bars or pubs and changes in rates of negative sequalae of alcohol consumption (eg, drunk driving fatalities and rates of cirrhosis).[29,30] Another relevant example was the association between banning sale of pesticides commonly used in suicides and rates of self-poisoning.[31] These types of studies are not always methodologically robust owing to potential unmeasured and unknown confounding.

However, a pre-post study design can be strengthened by including a control population not exposed to the intervention. Such studies permit an evaluation of background changes in rates of disorders over the relevant period. This approach assumes that the temporal trend of the control group provides a proxy for the trend that would have been observed in the exposed group in the absence of exposure. Thus, the difference in change of slope can be considered a causal effect size. Table 1 provides one such example: a study by Costello et al[22] on rural children aged 9 to 13 years in which receipt of additional income to families from the opening of a casino was associated with reduced rates of psychopathology in the children of these families but not matched families who did not receive the supplements. While the receipt of the additional funds was not random, the temporal association of the receipt of funds with the outcome and the lack of parallel changes in the control population argue in favor of a causal association.

Another compelling example of this method examines the association of homicide deaths by firearm and the adoption by Florida in late 2005 of the "Stand Your Ground" law[23] (Table 1; Figure 2D). The investigators showed, using monthly rates from 1999 to 2014, a clear temporal association between the rise in gun homicides after the law came into effect in Florida and no such change in the control states. These results support a causal interpretation of the association between the law and the subsequent rise of gun homicides.

© 2019 American Medical Association. All rights reserved.

Figure 2. Examples of Different Study Designs



A, A model for an instrumental variable analysis. The key feature is that the instrumental variable is associated with the risk factor and only associated with the disorder through the risk factor. B, Population A represents an illustration of the pre-post design, also called a regression discontinuity design, with longitudinal specification. The rates of the disorder change as a function of an intervention. Population B represents the control population that is not exposed to the intervention. C, Suicide rates from carbon monoxide (CO) poisoning and other non-CO related suicides. In 1963, the gas companies started decreasing the CO content of the gas. Reprinted with permission from Kreitman.[21] D, Data points represent monthly rates of homicide and homicide by firearms in Florida and comparison states (New York, New Jersey, Ohio, and Virginia) between 1999 and 2014. Florida is represented by orange data points and regression lines and the comparison states are represented by blue data points and regression lines. Gray-shaded areas depict the onset of Florida's "Stand Your Ground" law. Straight-hatched lines represent fitted estimates using a linear step change model. The curved lines represent fitted values for seasonally adjusted models. Reprinted with permission from Humphreys et al.[23]

Co-relative Designs

While the methods examined so far can potentially control all potential confounders, the next method, co-relative designs, makes more limited claims. Our thought experiment for ideal causal inference was to replicate individuals and study them in 2 worlds in which they were and were not exposed to the risk factor. The closest realistic approximation to that method is studying reared-together monozygotic twins discordant for risk-factor exposure. Such pairs share their genes at birth, develop in the same womb, are raised by the same parents, and are exposed during childhood and adolescences to the same physical, community, and school environments.

If a disorder is appreciably more common in the exposed vs unexposed members of such pairs, the intuition is that a causal inference may be made. However, while this method controls for all known and unknown genetic and shared-environmental confounders, it does not, as with classic or as-if randomization methods outlined previously, control for confounders affecting one of the twins. For example, imagine we examine impulsivity as a risk factor for drug abuse. Among monozygotic twins, the more impulsive twin has a considerably higher risk for drug abuse. This might appear to settle the issue of causation, but we noticed a small subset of these pairs were discordant for significant head injury in childhood. The injured twin consistently had higher impulsivity scores and higher rates of drug abuse. Now the potential for causal inference is limited because head injury is a confounder (Figure 1). Because head injury occurred to only 1 twin, it is not well controlled for in co-relative designs. In this and many similar examples, such injuries are rare and would likely produce only modest biases, but the principle holds.

Co-relative designs can be expanded by including other kinds of pairs of discordant relatives (eg, cousins and full siblings) and within-individual population estimates for the risk factor–disorder association. Using the simple rules of mendelian inheritance, examining such multiple groups can be used to estimate the association seen in discordant monozygotic pairs, which is often known imprecisely because of the rarity of such pairs.[20] Observing the expected decline in the association with increasing control for genetic and familial-environmental effects can increase confidence in the overall results and permit an assessment of the percentage of the population-based association that may be causal.

© 2019 American Medical Association. All rights reserved.

Statistical Models

A major advantage of natural experiments with plausible as-if random assignment of the risk factor is the simplicity of the statistical analysis. But such methods may not be available, in which case feasible approaches to causal inference include a range of statistical models applied to conventional observational studies. The most common analytic approach is multiple regression as applied either to cross-sectional or longitudinal data. To support causal inference from these models, investigators must identify and measure the important confounders and include them correctly in the statistical model. This requirement is difficult to meet with full confidence, but some studies have extensive sets of potential confounders to draw from, which ideally should be combined with sensitivity analyses for unmeasured confounding.[32,33] Additionally, the use of cross-sectional data introduces the additional problem of reverse causal association: we do not know the direction of the association and whether the risk factor preceded the disorder or the other way around.

A range of other methods have been proposed with the aim of improving the quality of causal inference. We examine 2 such methods: propensity score analysis (PSA)[34] and marginal structural models (MSM). The propensity score allows one to design and analyze an observational study mimicking key characteristics of an RCT. A propensity score is the probability of risk-factor exposure conditional on observed baseline characteristics. Thus, in a set of individuals, all of whom have the same propensity score, the distribution of observed baseline covariates will be the same between the exposed and unexposed study participants. The propensity score can then be used in various ways to remove the potential effects of known confounding variables: for example, by matching individuals with the same propensity but different exposure, or by including the propensity score as a covariate in a regression model. Regardless of the choice of method, conditioning on the propensity score may maximize statistical power, and on average result in measured baseline known covariates being balanced between exposure groups. Still, an important feature is to test whether this balance has been achieved.

Taylor et al[24] applied PSA to a series of RCTs of smoking reduction to examine whether quitting cigarette smoking improves mental health (Table 1). Uncorrected analyses showed a positive association between cessation of cigarette smoking and improved mental health. However, after applying PSA with a strong set of potential predictors of quitting, the statistical evidence for improved mental health disappeared.

Marginal structural models can be understood as a modification of the PSA that uses inverse-probability-of-treatment weighted estimators to balance those who were exposed and unexposed to the risk factor. In contrast to PSA, MSM can also be used when there exists a time-dependent risk factor for survival that is associated with subsequent treatment and when past treatment history is associated with subsequent risk factor level. However, both PSA and MSM and similar methods have to rely on potential confounders that are available in the data set and cannot, as with RCTs and natural experiments, control for unmeasured confounding.

Other Areas of Concern

Generalizability

In evaluating claims of causality in studies using methods for causal inference, the generalizability of results is sometimes given insufficient consideration. Causal claims, even if based on powerful designs, such as RCTs, may not be generalizable if they are performed among highly unrepresentative populations. This is also true for natural experiments. For example, should we assume that studies, such as that reviewed previously about the association between lottery winnings and mental health, help us to understand more generally the potential causal association between income and risk for psychiatric disorders? Clearly, there are concerns about generalizability because the nature of wealth arising from lottery winnings is different from how wealth is typically acquired. Despite their properties of randomization, studies on lottery winnings may be of limited scientific value in making casual inferences about associations between poverty or wealth and psychopathology.

Tests for Randomization

Claims about the quality of randomization in RCTs and particularly in natural experiments are not limited to conceptual analyses. Many data sets have a range of variables with which to test the quality of the randomization. For example, across a broad set of characteristics, a perfect randomization procedure should produce significant differences in the exposed and unexposed groups at an α level of .05 for approximately 5% of the variables examined. Results substantially in excess of that should raise suspicions about the quality of randomization.

Mechanisms

Western philosophy emphasizes 2 major approaches to the problems of causal inference: counterfactual (a comparison of the effect of an intervention in the real world with a hypothetical world in which the intervention did not occur) and mechanistic. As typical for epidemiologic and statistical approach to causal inference, all the models we have considered use a counter-factual framework. But how can mechanistic insights into pathways from risk factor to disease affect causal inference?

Several authors[5,35,36] have argued that in medicine, we need evidence from both counterfactual and mechanistic approaches to be confident about potential causal processes. While sympathetic to this position, we advocate a different view because the demonstration of casual mechanisms for psychiatric disorders is typically much more challenging than in other areas of medicine. We suggest that evidence for risk factor–disorder mechanisms strengthens the credibility of results obtained from counterfactual approaches but is not a necessary condition for causal inference. This is consistent with the Hill criteria for causality,[37] one of which was biologic plausibility (although in psychiatric epidemiology, the mechanisms are often psychological or social in nature). For example, Costello et al,[22] in further analyses, found that the association between extra income and childhood psychopathology was mediated through improved parental supervision. In our study on academic achievement and drug abuse misuse,[20] we reviewed other investigations showing that children who did poorly in school were prone to adopt a range of antisocial attitudes and behaviors including substance use and misuse. By contrast, a well-done RCT or natural experiment that provides evidence for a potential causal relationship for which no plausible biologic, psychological, or social mechanism exists should be greeted with some skepticism.

Level of Confounding

In most forms of causal inference from observational data (with the possible exception of those based on randomization), some re-

© 2019 American Medical Association. All rights reserved.

sidual confounding likely exists. The degree of confidence that should be placed in such results should relate to the likely magnitude of the confounding, the quality of the efforts to address the biases, and the possible presence of supporting information, especially those from methods with differing kinds of possible biases. Causally relevant information short of certainty remains valuable.

Triangulation

Given the limitations of any single method for assessing causal inference, confidence can be increased when evidence is found for potential causal relationships from several methods,[38] especially when they differ in their theoretical assumptions. Indeed, such evidence is stronger than replications of causal inference studies using the same method that may have hidden biases.[39] For example, a study in a twin population presented evidence for a causal association between dependent stressful life events and major depression using both co-twin control and PSA.[40] Our study of academic achievement and drug abuse risk[20] produced similar findings using instrumental variable and an expanded co-relative analysis. Triangulation is probably underused in efforts to clarify causal associations between risk factors and disorders in psychiatric epidemiology.

Reaching Conclusions

A range of approaches have been proposed to synthesize evidence for risk factor to disease causal inference, which range from qualitative summaries to formal Bayesian analyses[41] (see chapter 8 of Samet and Bodurow[42] for a review). An Institute of Medicine committee identified 4 categories, with definitions, for "the strength of the overall evidence for or against a causal relationship from exposure to disease" (**Table 2**).[42] Based on epidemiologic data only, to meet the highest level of "sufficient" evidence requires "replicated and consistent evidence of a causal association: that is, evidence of an association from several high-quality epidemiologic studies that cannot be explained by plausible noncausal alternatives."[42(p189)] While beyond the scope of our review, several classical risk factors for psychiatric disorders would probably meet this criterion, while more would meet the less rigorous "equipoise and greater" (Table 2).

Table 2. Proposed Categories for the Level of Evidence for Causation (Chapter 8)[42]

Category	Definition
Sufficient	The evidence is sufficient to conclude that a causal relationship exists
Equipoise and greater	The evidence is sufficient to conclude that a causal relationship is at least as likely as not, but not sufficient to conclude that a causal relationship exists
Less than equipoise	The evidence is not sufficient to conclude that a causal relationship is at least as likely as not or is not sufficient to make a scientifically informed judgment
Against	The evidence suggests the lack of a causal relationship

Conclusions

Causal inference is important because it may inform prevention efforts and etiologic model building in a more useful way than statistical associations. A diversity of methods is available to the epidemiologist attempting to gain insight into the potential causal nature of an association between putative risk factors and disorders. Herein, we reviewed a number of the major approaches and their relative strengths and limitations. There are several methods that we have not discussed with increasing popularity such as agent-based models[43] and mendelian randomization.[26,44] In closing, we argue against a common view that causation only can be claimed from RCTs but that no other analytical method can provide useful causal inferences. Notably, as the Hill criteria[37] long ago made clear, no single study (even an RCT) can provide unshakable evidence for causation, especially in the generalized populations that are usually of maximal interest. Each of the various methods has potential limitations. We should avoid the extremes of overzealous causal claims, especially from single studies using a single method, but we should also avoid the cynical view that useful potential causal information is unattainable with the methods herein reviewed except RCTs. In making causal claims, care and self-critical circumspection is needed. Misattribution of causality in matters of public health is not just an academic question because incorrect claims of causality can cause harm.[45] Even if uncertain, good-quality information about the plausibility of potential causal claims is important for research and public health.

ARTICLE INFORMATION

Published Online: December 11, 2019. doi:10.1001/jamapsychiatry.2019.3758

Accepted for Publication: September 26, 2019.

Conflict of Interest Disclosures: None reported.

Funding/Support: This project was supported by grants R01DA030005 and R01AA023534 from the National Institutes of Health, the Swedish Research Council (K2012-70X-15428-08-3), the Swedish Research Council for Health, Working Life and Welfare, the Swedish Research Council (2012-2378; 2014-10134), and FORTE (2014-0804) as well as ALF funding from Region Skåne awarded.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: Sean Clouston, PhD, Department of Family, Population, and Preventive Medicine, Program in Public Health, Stony Brook University Health Sciences Center, Stony Brook, New York, provided helpful comments on an early version of this manuscript. No compensation from a funding source was received.

REFERENCES

1. Woodward J. *Making Things Happen*. New York: Oxford University Press; 2003.

2. Glass TA, Goodman SN, Hernán MA, Samet JM. Causal inference in public health. *Annu Rev Public Health*. 2013;34:61-75. doi:10.1146/annurev-publhealth-031811-124606

3. Susser M. *Causal Thinking in the Health Sciences*. New York, NY: Oxford University Press; 1973.

4. Dunning T. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge, UK: Cambridge University Press; 2012. doi:10.1017/CBO9781139084444

5. Gillies D. *Causality, Probability, and Medicine*. London, England: Routledge: Taylor & Francis Group; 2019.

6. Holland PW. Statistics and causal inference. *J Am Stat Assoc*. 1986;81(396):945-960. doi:10.1080/01621459.1986.10478354

7. Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge, England: Cambridge University Press; 2013.

8. Kenny DA. *Correlation and Causality*. New York, NY: Wiley-Interscience; 1979.

9. Hernán MA, Robins JM. *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC; 2019.

10. Broglio K. Randomization in clinical trials: permuted blocks and stratification. *JAMA*. 2018;319(21):2223-2224. doi:10.1001/jama.2018.6360

© 2019 American Medical Association. All rights reserved.

11. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2-21. doi:10.1016/j.socscimed.2017.12.005

12. Frieden TR. Evidence for health decision making: beyond randomized, controlled trials. *N Engl J Med*. 2017;377(5):465-475. doi:10.1056/NEJMra1614394

13. Hohmann E, Brand JC, Rossi MJ, Lubowitz JH. Expert opinion is necessary: Delphi panel methodology facilitates a scientific approach to consensus. *Arthroscopy*. 2018;34(2):349-351. doi:10.1016/j.arthro.2017.11.022

14. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. *Am J Prev Med*. 2007;33(2):155-161. doi:10.1016/j.amepre.2007.04.007

15. Ravallion M. Should the Randomistas (Continue to) Rule? CDG Working Paper 492. 2018. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/should-randomistas-continue-rule. Accessed August 10, 2019.

16. Rutter M. Proceeding from observed correlation to causal inference: the use of natural experiments. *Perspect Psychol Sci*. 2007;2(4):377-395. doi:10.1111/j.1745-6916.2007.00050.x

17. Apouey B, Clark AE. Winning big but feeling no better? the effect of lottery prizes on physical and mental health. *Health Econ*. 2015;24(5):516-538. doi:10.1002/hec.3035

18. Snow J. *On The Mode of Communication of Cholera. London, UK: John Churchill*. London, England: New Burlington Street; 1855.

19. Wang PS, Schneeweiss S, Avorn J, et al. Risk of death in elderly users of conventional vs. atypical antipsychotic medications. *N Engl J Med*. 2005;353(22):2335-2341. doi:10.1056/NEJMoa052827

20. Kendler KS, Ohlsson H, Fagan AA, Lichtenstein P, Sundquist J, Sundquist K. Academic achievement and drug abuse risk assessed using instrumental variable analysis and co-relative designs. *JAMA Psychiatry*. 2018;75(11):1182-1188. doi:10.1001/jamapsychiatry.2018.2337

21. Kreitman N. The coal gas story: United Kingdom suicide rates, 1960-71. *Br J Prev Soc Med*. 1976;30(2):86-93. doi:10.1136/jech.30.2.86

22. Costello EJ, Compton SN, Keeler G, Angold A. Relationships between poverty and psychopathology: a natural experiment. *JAMA*. 2003;290(15):2023-2029. doi:10.1001/jama.290.15.2023

23. Humphreys DK, Gasparrini A, Wiebe DJ. Evaluating the impact of Florida's "Stand Your Ground" self-defense law on homicide and suicide by firearm: an interrupted time series study. *JAMA Intern Med*. 2017;177(1):44-50. doi:10.1001/jamainternmed.2016.6811

24. Taylor G, Girling A, McNeill A, Aveyard P. Does smoking cessation result in improved mental health? a comparison of regression modelling and propensity score matching. *BMJ Open*. 2015;5(10):e008774. doi:10.1136/bmjopen-2015-008774

25. Maciejewski ML, Brookhart MA. Using instrumental variables to address bias from unobserved confounders. *JAMA*. 2019;321(21):2124-2125. doi:10.1001/jama.2019.5646

26. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. *JAMA*. 2017;318(19):1925-1926. doi:10.1001/jama.2017.17219

27. Clouston SA, Denier N. Mental retirement and health selection: analyses from the US Health and Retirement Study. *Soc Sci Med*. 2017;178:78-86. doi:10.1016/j.socscimed.2017.01.019

28. Daigle MS. Suicide prevention through means restriction: assessing the risk of substitution: A critical review and synthesis. *Accid Anal Prev*. 2005;37(4):625-632. doi:10.1016/j.aap.2005.03.004

29. Babor T, Caetano R, Casswell S, et al. *Alcohol: No Ordinary Commodity: Research and Public Policy*. 2nd ed. Oxford, England: Oxford University Press; 2010. doi:10.1093/acprof:oso/9780199551149.001.0001

30. Thern E, Carslake D, Davey Smith G, Tynelius P, Rasmussen F. The effect of increased alcohol availability on alcohol-related health problems up to the age of 42 among children exposed in utero: a natural experiment. *Alcohol Alcohol*. 2018;53(1):104-111. doi:10.1093/alcalc/agx069

31. Gunnell D, Fernando R, Hewagama M, Priyangika WD, Konradsen F, Eddleston M. The impact of pesticide regulations on suicide in Sri Lanka. *Int J Epidemiol*. 2007;36(6):1235-1242. doi:10.1093/ije/dym164

32. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med*. 2017;167(4):268-274. doi:10.7326/M16-2607

33. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-value to assess the potential effect of unmeasured confounding in observational studies. *JAMA*. 2019;321(6):602-603. doi:10.1001/jama.2018.21554

34. Haukoos JS, Lewis RJ. The propensity score. *JAMA*. 2015;314(15):1637-1638. doi:10.1001/jama.2015.13480

35. Canali S. Evaluating evidential pluralism in epidemiology: mechanistic evidence in exposome research. *Hist Philos Life Sci*. 2019;41(1):4. doi:10.1007/s40656-019-0241-6

36. Russo F, Williamson J. Interpreting causality in the health sciences. *Philos Sci*. 2007;21(2):157-170. doi:10.1080/02698590701498084

37. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58(4):295-300. doi:10.1177/003591576505800503

38. Munafò MR, Davey Smith G. Robust research needs many lines of evidence. *Nature*. 2018;553(7689):399-401. doi:10.1038/d41586-018-01023-3

39. Bonovas S, Filioussi K, Flordellis CS, Sitaras NM. Statins and the risk of colorectal cancer: a meta-analysis of 18 studies involving more than 1.5 million patients. *J Clin Oncol*. 2007;25(23):3462-3468. doi:10.1200/JCO.2007.10.8936

40. Kendler KS, Gardner CO. Dependent stressful life events and prior depressive episodes in the prediction of major depression: the problem of causal inference in psychiatric epidemiology. *Arch Gen Psychiatry*. 2010;67(11):1120-1127. doi:10.1001/archgenpsychiatry.2010.136

41. McGlothlin AE, Viele K. Bayesian hierarchical models. *JAMA*. 2018;320(22):2365-2366. doi:10.1001/jama.2018.17977

42. Samet JM, Bodurow CC. *Improving The Presumptive Disability Decision-Making Process for Veterans*. Washington, DC: The National Academies Press: Institute of Medicine of the National Academies; 2008.

43. Tracy M, Cerdá M, Keyes KM. Agent-based modeling in public health: current applications and future directions. *Annu Rev Public Health*. 2018;39:77-94. doi:10.1146/annurev-publhealth-040617-014317

44. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658-665. doi:10.1002/gepi.21758

45. Rossouw JE, Anderson GL, Prentice RL, et al; Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288(3):321-333. doi:10.1001/jama.288.3.321

© 2019 American Medical Association. All rights reserved.