



Published in final edited form as:

JAMA Psychiatry. 2017 June 01; 74(6): 561–562. doi:10.1001/jamapsychiatry.2017.0502.

Causal Inference in Psychiatric Epidemiology

Kenneth S. Kendler, MD

Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond; Department of Psychiatry, Virginia Commonwealth University, Richmond

There is no question more fundamental for observational epidemiology than that of causal inference. When, for practical or ethical reasons, experiments are impossible, how may we gain insight into the causal relationship between exposures and outcomes? This is the key question that Quinn et al¹ seek to answer: does maternal smoking during pregnancy (SDP) cause offspring serious mental illness (SMI)?

The problem of causation has a huge, fascinating, and occasionally bewildering literature with major contributions from philosophy, statistics, and epidemiology (see Susser,² Woodward,³ and Illari and Russo⁴ for helpful introductions). Our focus here will be on the problem of confounding.

As depicted in Figure 1A, we want to understand how much of the association between the exposure (here SDP) and the outcome (SMI) is the result of the direct causal path vs a result of the two groups of possible confounders: those measured and those not measured. The functional definition of a confounder is an external variable that predicts both the exposure and the outcome, and thus causes the two to correlate.

The most common approach to the problem of confounding is multiple regression, where the outcome is predicted from the exposure and measured confounders. The hope is that the resulting exposure-outcome relationship will reflect causal effects. But this approach suffers from a major problem. Have you identified all the possible confounders? And, even if you know them, are they well measured in your sample?

Figure 1B depicts a co-relative approach to the problems of confounders, used by Quinn et al.¹ Here confounders are divided into familial and nonfamilial. As a rough definition, think of “familial” as confounding variables that are substantially correlated in siblings who grew up together. A strength of this design is that a large majority of human behavioral traits are correlated in relatives, often substantially.⁵ Think of nonfamilial confounders as experiences that make siblings different from one another in both risk for the exposure and the outcome. A major difference between Figure A and Figure B is that the familial confounders do not need to be measured nor does the researcher need to even be aware of their existence.

Corresponding Author: Kenneth S. Kendler, MD, Virginia Institute for Psychiatric and Behavioral Genetics of VCU, Box 980126, Richmond, VA 23298-0126 (kenneth.kendler@vcuhealth.org).

Conflict of Interest Disclosures: None reported.

Additional Contributions: Helpful comments were provided on an earlier version of this manuscript by Henrik Ohlsson, PhD (Center for Primary Health Care Research, Lund University, Malmö, Sweden) and Charles Gardner, PhD (Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond). No compensation was received.

Comparing the rates of the outcome in siblings who were vs were not exposed controls for the familial components of *all* confounders. Co-relative designs resemble randomized controlled trials in that they account for the effect of both known and unknown confounders.

But there is a problem. The co-relative design does not work for nonfamilial confounders—those unique to individuals. An example may clarify this point:

Consider a sibling pair discordant for impulsivity (the exposure) where the impulsive sibling has poor school performance while the unimpulsive sib does not. This pair seems to provide evidence for a causal effect of impulsivity on school achievement. But, what if the affected sibling had a serious head trauma at age 8 with subsequent impulsivity and poor school performance? Maybe the relationship is not causal but due to a nonfamilial confounder—head trauma.

So, while very helpful, the co-relative design is no panacea for all problems of causal inference. Two more points about this design are noteworthy. First, it only controls fully for genetic confounding in monozygotic (MZ) twin pairs discordant for exposure. Finding enough such pairs, even in Scandinavian registries, can be difficult. For intrauterine exposures like SDP, furthermore, twins are always concordant. However, we have relatives that share different degrees of environmental exposures and genes such as cousins, half-siblings, full siblings, and MZ twins. These can be used to create a “dose-response” curve of increasing control for familial confounders. With such data, it is possible to fit a model to estimate results for discordant MZ twins from the other more common relative pairs.⁶

Second, the interpretation of the co-relative design is somewhat asymmetric. When, as in Quinn et al,¹ the exposure outcome relationship is substantially attenuated in co-relative pairs, one can be relatively confident that familial confounders are at work, so that causal processes alone cannot explain the findings. However, when the association does not attenuate in co-relative designs, more caution is needed. Such findings suggest that causal processes explain most of the exposure-outcome relationship, but could nonfamilial confounders be making a major contribution?

Many other methods seek to clarify causal effects in observational data. They can be usefully divided into two major groups. The first are statistical methods, for example propensity score matching⁷ and marginal structural models.⁸ Both methods work best with rich sets of predictors of exposure. Propensity score is conceptually elegant in its selection, from available data, of pairs of individuals with equal propensities to exposure where one has and the other has not been actually exposed. Marginal structural models are best applied to longitudinal data on exposures and outcomes.

The second are true natural experiments, the best of which provide powerful instrumental variables. One excellent example is the study of Costello et al⁹ examining the impact of increased family income on child psychopathology. It can be particularly instructive to compare the results of co-relative and statistical approaches to causal inference in the same data sets.¹⁰

The Quinn et al¹ results are relatively convincing. The SDP-SMI association is substantially attenuated with measured confounders and discordant cousin pairs. The attenuation is even greater in sibling pairs and loses statistical significance. They also examined measured confounders that might reflect nonfamilial effects within their co-relative design. In a lovely move, they used the co-relative design to show that the impact of SDP on two obstetric outcomes (small for gestational age and preterm birth) attenuates much less in relative pairs than SMI and so is likely a true consequence of SDP.

I close with these cautions. With observational data, we can never be certain about causal processes. We can only seek for increased confidence that causal effects are likely present. It is this confidence that can help guide planned prevention and intervention efforts. However, this issue has a flip side. Finding associations in observational data are too easy. Researchers who report such results are obliged to avoid causal language and wherever possible, use available methods to provide some insight into the possible causal relationship between their exposure and their outcome.

Acknowledgments

Funding/Support: This was supported in part by NIH grants R01DA030005 and R01AA023534

Role of the Funder/Sponsor: The funding sources had no role in the preparation, review, or approval of the manuscript, or the decision to submit the manuscript for publication.

References

1. Quinn, PD., Rickert, ME., Weibull, CE., et al. Association between maternal smoking during pregnancy and severe mental illness in offspring [published online May 3, 2017]. *JAMA Psychiatry*. doi:10.1001/jamapsychiatry.2017.0456
2. Susser M. What is a cause and how do we know one? a grammar for pragmatic epidemiology. *Am J Epidemiol*. 1991; 133(7):635–648. [PubMed: 2018019]
3. Woodward, J. *Making Things Happen*. New York: Oxford University Press; 2003.
4. Illari, P., Russo, F. *Causality: Philosophical Theory Meets Scientific Practice*. Oxford, UK: Oxford University Press; 2014.
5. Polderman TJ, Benyamin B, de Leeuw CA, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. 2015; 47(7):702–709. [PubMed: 25985137]
6. Kendler KS, Ohlsson H, Sundquist J, Sundquist K. Alcohol use disorder and mortality across the lifespan: a longitudinal cohort and co-relative analysis. *JAMA Psychiatry*. 2016; 73(6):575–581. [PubMed: 27097014]
7. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiol Drug Saf*. 2004; 13(12): 841–853. [PubMed: 15386709]
8. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11(5):550–560. [PubMed: 10955408]
9. Costello EJ, Compton SN, Keeler G, Angold A. Relationships between poverty and psychopathology: a natural experiment. *JAMA*. 2003; 290(15):2023–2029. [PubMed: 14559956]
10. 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. [PubMed: 21041613]

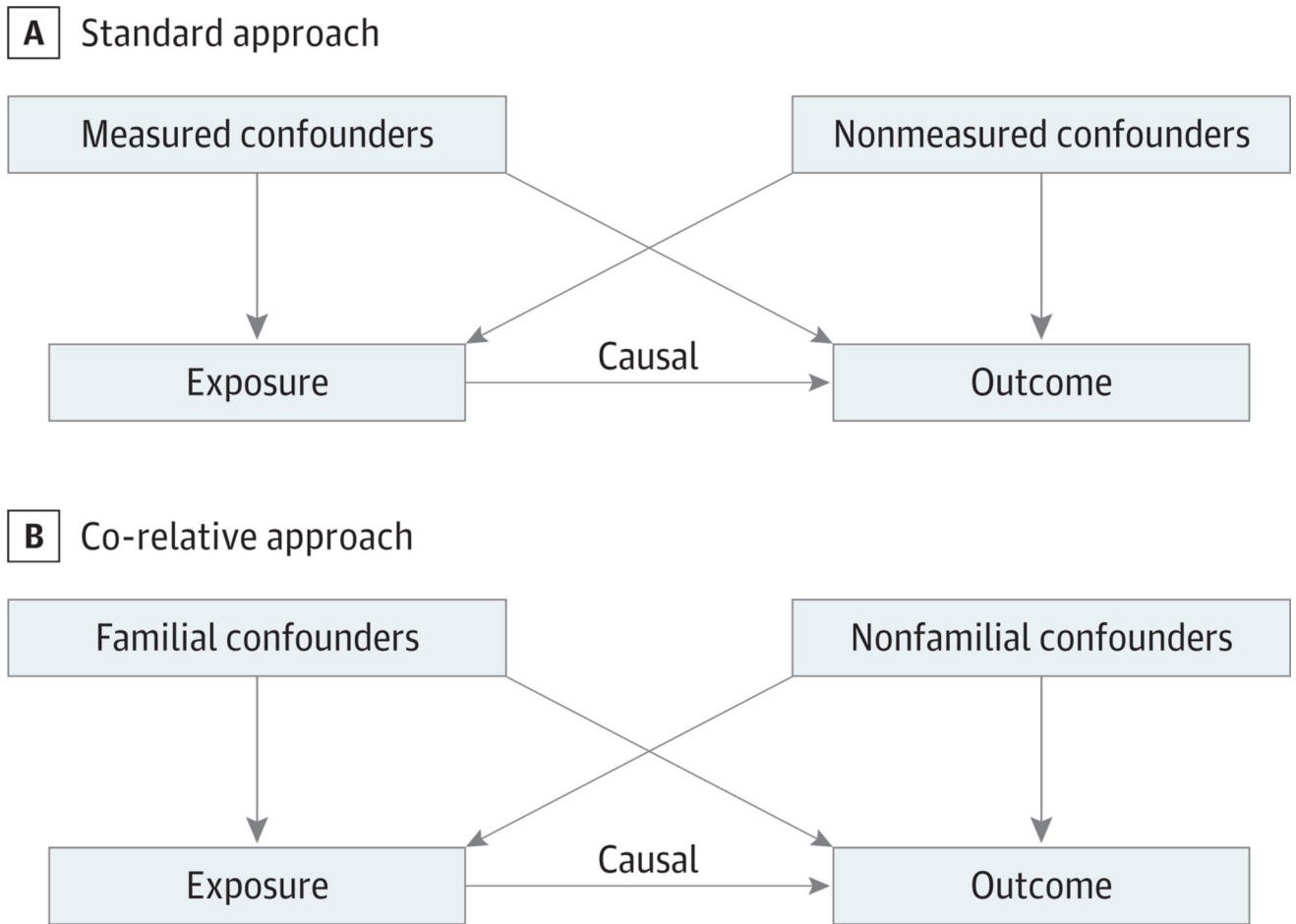


Figure 1. Approaches to Causal Inference

A, The standard approach to causal inference between an exposure (or risk factor) and outcome using multiple regression. Such methods can only include measured confounders. However, the impact of unmeasured confounders can bias upward the estimate of the causal relationship between the exposure and the outcome. B, A co-relative approach to causal inference between an exposure (or risk factor) and outcome. This method controls for all familial confounders whether measured or unmeasured. However, the impact of nonfamilial confounders can bias upward the estimate of the causal relationship between the exposure and the outcome.