

Making Causal Inferences From Observational Databases?

T. Michael Kashner, Ph.D., J.D.

Developing statistical models to make causal inferences from observational data offers important challenges for psychiatric researchers. Unlike variables that merely covary in a data set, A is said to cause B if the following conditions are met: 1) *plausibility*: A impacts B is consistent with accepted theory; 2) *temporal precedence*: A precedes B in time; 3) *simultaneous effects*: the effect of A on B can be distinguished from the effect, if any, of B on A ; and 4) *exogeneity*: A affects B after all confounding factors C are held constant. Whether known or unknown to the researcher, confounding factors are associated with both A and B and, if left unaccounted for, will contribute to the observed correlation between A and B .

In this issue of the *Journal*, Blanco et al. (1) estimated the association between pain severity and prescription opioid use disorder in a representative sample of noninstitutionalized adults. They found that patients with painful conditions were at higher risk of opioid use disorder, and they conclude that “careful monitoring and consideration of nonopioid alternative treatments is warranted.”

The authors addressed the above causality conditions, including simultaneous effects, by applying structural equation modeling to the survey’s baseline (wave 1) and follow-up (wave 2) data structure and selecting confounders from among “several background demographic and clinical characteristics.” This approach works when the cause (in this case, pain) is outside the control of the patient and a well-designed database includes all critically important confounding variables that affect outcomes. It is, however, worth noting what happens if important unobserved confounders are present.

Following Heckman and Navarro-Lozano (2) and refined by Lu and White (3), control functions are designed to account for both observable and unobservable confounders. Here, we can let y be the outcome opioid use disorder and D be the variable of interest, pain severity, with Z as the vector of observable covariates of outcome that are available in the data set and U as the vector of remaining unobservable covariates that are not available in the data set. If the true data-generating process of outcome is linear, then:

$$y = \mu_0 + D\beta_0 + Z'\alpha_0 + U'\gamma_0 + \epsilon_0, \quad [1]$$

where the coefficient to D , β_0 , is the true pain effect size and the parameter of interest. Similarly, α_0 and γ_0 are the true effects of Z and U , respectively, μ_0 is the intercept parameter, and ϵ_0 is an independent and identically distributed random error.

To estimate β_0 , suppose we model the contribution of the unobserved covariates to outcome as

$$[U'\gamma_0] = \delta_c + D\delta_D + Z'\delta_Z + W'_0\delta_{W_0} + W'_1\delta_{W_1} + \epsilon_1, \quad [2]$$

where W_0 are proxies for the missing opioid use disorder covariates and W_1 are additional variables representing drivers of pain (D) and proxies of drivers of pain, with model parameters δ_i for $i=c, D, Z, W_0$, and W_1 . ϵ_1 is an independent and identically distributed random error. Substituting equation 2 into equation 1 yields:

$$y = (\mu_0 + \delta_c) + D(\beta_0 + \delta_D) + Z'(\alpha_0 + \delta_Z) + W'_0\delta_{W_0} + W'_1\delta_{W_1} + \epsilon_0 + \epsilon_1. \quad [3]$$

Note that the parameter coefficient to D equals the true effect β_0 plus confounding bias δ_D . The strategy is to deliberately misspecify equation 2 by entering additional variables W_1 so that pain (D) will no longer contribute to the total impact of the unobserved covariates in equation 2. One can test for bias by taking a transform (squared) of the estimated residuals in equation 3, regressing on the right-hand side of equation 2, and testing directly for $(\delta_D \neq 0)$. By re-estimating equation 3 for different sets of qualifying W_1 variables, one can test whether the resulting effect size estimates vary—i.e., are robust. Since the focus on adding variables is to compute an unbiased pain effect, estimates of the coefficients to Z variables will likely remain biased ($\delta_Z \neq 0$), leaving the estimated effect of Z on opioid use disorder risk uninterpretable.

There are other strategies. For instance, two-stage residual inclusion replaces W_1 with the estimated residuals to regressing pain on variables that are correlated with pain, do not drive opioid use disorder, and are uncorrelated with U and ϵ_0 (4).

If present, correcting for unobserved confounders can make a difference. In our study of over 12,500 Veterans Affairs physician residents (5), we found that an association between psychological safe learning environments and physician resident satisfaction with clinical learning after adjusting for 16 factors fell by 50% when the estimate was further adjusted for unobserved confounders using a two-stage residual inclusion method.

One problem with control functions is that effect size estimates for the other covariates often remain biased and thus are uninterpretable. We have investigated using an approach designed to identify the true data-generating process and then estimate its parameters. “Best approximating models” (6, 7) begin with a stochastic/exhaustive search of all possible

models constructed from a list of plausible covariates, plus nonlinear transforms and interaction terms. Each model is estimated to account for over-fitting and multicollinearity. The best-fitting model is selected from among those that do not test positive for model misspecification. The key here is to test whether the researcher's final model is correctly specified to represent the true data-generating process (8, 9). As described above, models that can predict may not necessarily reflect the true data-generating process. For instance, a misspecified candidate model that differed by an incorrect transformation and irrelevant predictor from a data-generating process third-order power series was found to be observationally equivalent, yet it tested positive for model misspecification (8, 9). In simulation experiments in head-to-head comparisons with propensity score matching (which does not account for unobserved confounders), best approximating models were shown to differ by less than 10% of the true effect size in the presence of unobserved confounders, compared with estimates from propensity score matching that fell between 75% below and 50% above the true effect size (10, 11).

While Blanco et al. are to be applauded for using advanced analytics to compute true associations from observational data sets, investigators should be encouraged to consider more principled and systematic approaches in selecting their covariates, followed by testing their performance to see whether the variable of interest is conditionally exogenous, or the final model representing the data-generating process is correctly specified.

AUTHOR AND ARTICLE INFORMATION

From the Department of Medicine, Loma Linda University Medical School, Loma Linda, Calif.; the Department of Psychiatry, University of Texas Southwestern Medical Center at Dallas; and the Office of Academic Affiliations, Department of Veterans Affairs, Washington, D.C.

Address correspondence to Dr. Kashner (michael.kashner@va.gov).

The author reports no financial relationships with commercial interests.

Accepted September 2016.

Am J Psychiatry 2016; 173:1161–1162; doi: 10.1176/appi.ajp.2016.16091091

REFERENCES

1. Blanco C, Wall MM, Okuda M, et al: Pain as a predictor of opioid use disorder in a nationally representative sample. *Am J Psychiatry* 2016; 173:1189–1195
2. Heckman J, Navarro-Lozano S: Using matching, instrumental variables, and control functions to estimate economic choice models. *Rev Econ Stat* 2004; 86:30–57
3. Lu X, White H: Robustness checks and robustness tests in applied economics. *J Econom* 2014; 178:194–206
4. Terza JV, Basu A, Rathouz PJ: Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J Health Econ* 2008; 27:531–543
5. Torralba KD, Loo LK, Byrne JM, et al: Does psychological safety impact the clinical learning environment for physician residents? Results from VA's Learners' Perceptions Survey. *J Grad Med Educ* (Epub ahead of print, Oct 6, 2016)
6. Westover AN, Kashner TM, Winhusen TM, et al: A systematic approach to subgroup analyses in a smoking cessation trial. *Am J Drug Alcohol Abuse* 2015; 41:498–507
7. Henley SS, Kashner TM, Golden RM, et al: Response to letter regarding "A systematic approach to subgroup analyses in a smoking cessation trial". *Am J Drug Alcohol Abuse* 2016; 42:112–113
8. Golden RM, Henley SS, White H Jr, et al: New directions in information matrix testing: Eigenspectrum tests, in *Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions* Essays in Honour of Halbert L. White, Jr. Edited by Swanson NR, Chen X. New York, Springer Science and Business Media, 2013, pp 145–177
9. Golden RM, Henley SS, White H, Kashner TM. Generalized information matrix tests for detecting model misspecification. *Econometrics* (in press).
10. Kashner TM, Chen GJ, Golden RM, et al: Recommendations from HSR&D's Panel on Statistics and Analytics on VHA Datasets. Presented at the 2015 Health Services Research and Development/Quality Enhancement Research Initiative, Philadelphia, July 8–10, 2015
11. Kashner TM, Henley SS, Golden RM: New strategies to solve analytic challenges in HSR. Workshop presented at the Annual Research Meeting of Academy Health, Boston, June 25–28, 2016