Published in final edited form as:

Epidemiology. 2014 March ; 25(2): 279–281. doi:10.1097/EDE.0000000000000068.

Balancing Automated Procedures for Confounding Control with Background Knowledge

Richard Wyss and Til Stürmer

In the publication by Patorno *et al.*¹ found in this issue of *Epidemiology*, the authors illustrate the importance of using subject matter knowledge to complement the automated high-dimensional propensity score (hdPS) algorithm when controlling for confounding in studies based on claims data with few exposed outcomes. The topic of variable selection for PS models in settings involving large numbers of potential confounders has received considerable attention in recent years. This interest is in part due to the uncertainty in determining what role automated procedures should play in the variable selection process. With large healthcare databases becoming increasingly used in epidemiology,^{2–4} automated procedures can be beneficial in selecting potential confounders that are unknown to the investigator.^{5–7} Further, the application of automated procedures is likely to expand as safety surveillance receives more attention as part of the Food and Drug Administration's Sentinel Initiative.⁸ In these settings, automated procedures, such as the hdPS, can increase the speed and efficiency of active surveillance.⁷ With an increasing need for automated methods for confounding control in these areas of epidemiologic research, the question becomes: how should investigators balance automated procedures with the use of subject matter knowledge?

Although automated procedures can be beneficial in identifying empirical associations among large numbers of covariates, empirical associations and probability distributions by themselves are not sufficient to determine causal relations.^{9–13} In a recent commentary, Pearl¹³ explains that probability distributions of observed variables cannot completely characterize causal relationships and that “every exercise in causal inference must commence with some extra knowledge that cannot be expressed in probability alone.” On the topic of variable selection, many authors have argued that the identification of confounding variables should be grounded in prior substantive knowledge.^{9, 10, 11} In discussing this issue, Hernan *et al.*¹¹ emphasize that “causal inference from observational data requires prior causal assumptions or beliefs, which must be derived from subject matter knowledge, not from statistical associations detected in the data.”

While it is easy to acknowledge the theoretical limitations of using empirical associations to identify causal relations, the practical consequences of these limitations are less clear. For example, a potential obstacle for automated variable selection procedures is the possibility for an increase in bias amplification caused by controlling for instrumental variables and

Author for correspondence: Til Stürmer, MD, MPH, PhD, Department of Epidemiology, UNC Gillings School of Global Public Health, University of North Carolina at Chapel Hill, McGavran-Greenberg, CB # 7435, Chapel Hill, NC 27599-7435, Phone: +1 919 966 7433, Fax: +1 919 966 2089, til.sturmer@post.harvard.edu.

collider-stratification bias (e.g., M-bias). Although the negative effects of these biases are obvious in theory, the impact of bias amplification and collider-stratification bias in practice is elusive. Recent simulation studies have examined the magnitude of these biases in several practical settings and have shown that such increases in bias are generally small compared with the bias resulting from the exclusion of confounding variables.^{14, 15} Based on their results, Myers *et al.*¹⁴ and Liu *et al.*¹⁵ recommend that controlling for confounding should take precedence over avoiding bias amplification¹⁴ or M-bias.¹⁵ While we agree with this recommendation, it is important to recognize that there do exist scenarios where collider-stratification bias and, in particular, bias amplification can be substantial.

In his commentary on the Meyers *et al.* study, Pearl¹⁶ describes theoretical situations where the cumulative effect of conditioning on multiple variables that have strong effects on treatment but weak effects on the outcome (i.e., near-instruments) can result in bias amplification that is more pronounced than the bias reduction obtained from controlling those variables. More specifically, Pearl¹⁶ states that “the cumulative effect of sequential conditioning has a built-in slant towards bias amplification as compared with confounding reduction; the latter is tempered by sign cancellations, the former is not.” Published empirical examples that demonstrate bias amplification or collider-stratification bias are rare but do exist. One such study by Patrick *et al.*¹⁷ reported that the inclusion of a glaucoma diagnosis covariate (a strong predictor of receiving the comparator drug; anti-glaucoma medication) in the propensity score resulted in a substantial increase in bias when estimating the effect of statins on mortality and hip fracture.

Although bias amplification and collider-stratification bias can be substantial in specific settings, there is little empirical evidence of these obstacles having a significant impact on the automated hdPS algorithm. Several studies have demonstrated that the hdPS, when used to complement investigator-specified covariate adjustment, often improves confounding control and performs no worse than investigator-specified approaches used by themselves.^{5-7, 18} In discussing the hdPS for confounding adjustment in large healthcare databases, Rassen and Schneeweiss⁷ conclude that “any confounding bias will likely be greater in magnitude than collider bias” and “an automated confounding adjustment system that selects a large number of covariates, even with somewhat imperfect variable selection, should improve study validity far more than it will harm it.” While we tend to agree, it remains that investigators should be aware of the potential limitations that bias amplification and collider-stratification bias present and the potential for these biases to impact hdPS performance in some circumstances. Researchers should further be aware of the possibility for additional limitations to arise as the hdPS is applied in settings where its performance has not been well established. For example, recent studies have described situations involving few exposed events where the performance of the automated hdPS algorithm may be limited.^{6, 7, 18}

The theoretical limitations outlined above dictate that automated procedures for confounding control are not optimal in every situation. Indeed, there have been published examples where the hdPS, when used as a replacement for investigator-specified covariate adjustment, increased bias compared with covariate adjustment procedures that incorporate expert knowledge.^{1, 19} In the study by Toh *et al.*¹⁹ based on a UK electronic medical record

database, one of the key points reported was “compared with adjustment for investigator-identified variables only, adjustment using the hd-PS algorithm (with only age and sex included *a priori*) was closer to the unadjusted estimate.” The study by Patorno *et al.*¹ found in this issue of *Epidemiology* is the first to empirically demonstrate within claims data that the hdPS can perform worse, as compared with investigator-specified covariate adjustment, when used in the absence of a pre-specified set of investigator-selected variables. Their results highlight that the hdPS can be especially vulnerable in settings with few exposed outcomes. These findings are valuable contributions to the literature as they help us understand in which specific settings the theoretical limitations of using automated procedures in the absence of investigator input can have practical consequences and should thus be avoided.

As the need for automated procedures for confounding control grows, researchers will continue to apply the hdPS in novel studies and new areas where its performance is not well established. It is in these settings that it is important to recognize that there do exist situations in both theory and practice where the hdPS, and automated procedures in general, can perform poorly when used in the absence of investigator input. We therefore emphasize the importance of using substantive knowledge to obtain an understanding of the data and the underlying causal structure before applying automated procedures for confounding control.¹⁰ While the hdPS is a valuable tool in helping researchers control for large numbers of potential confounders, we believe that automated procedures should not replace background knowledge, but should be used to complement investigator input when controlling for confounding as emphasized by Patorno *et al.*¹

References

1. Patorno E, Glynn RJ, Hernandez-Diaz S, Liu J, Schneeweiss S. Covariate selection and implementation of propensity-score-based confounding adjustment in studies with many covariates and few outcomes: Anticonvulsants and cardiovascular risk. *Epidemiology*. In Press.
2. Ray WA. Population-based studies of adverse drug effects. *N Engl J Med*. 2003; 365:475–81.
3. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res*. 2009; 18:67–80. [PubMed: 18562398]
4. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005; 58:323–337. [PubMed: 15862718]
5. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009; 20(4):512–522. [PubMed: 19487948]
6. Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am J Epidemiol*. 2011; 173(12): 1404–1413. [PubMed: 21602301]
7. Rassen JA, Schneeweiss S. Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. *Pharmacoepidemiol Drug Saf*. 2012; 21(S1):41–29. [PubMed: 22262592]
8. Platt R, Wilson M, Chan A, Benner JS, Marchibroda J, McClellan M. The new sentinel network-improving the evidence of medical-product safety. *N Engl J Med*. 2009; 361:645–647. [PubMed: 19635947]
9. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999; 10:37–48. [PubMed: 9888278]

10. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology*. 2001; 12:313–20. [PubMed: 11338312]
11. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *Am J of Epidemiol*. 2002; 155(2):176–184. [PubMed: 11790682]
12. Pearl, J. *Causality: models, reasoning, and inference*. 2. Cambridge University Press; New York: 2009.
13. Pearl J. Comment on ‘Causal inference, probability theory, and graphical insights’ by Stuart G. Baker. *Stat Med*. 2013; 32:4331–4333.
14. Myers JA, Rassen JA, Gagne JJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol*. 2011; 174(11):1213–1222. [PubMed: 22025356]
15. Liu W, Brookhart MA, Schneeweiss S, Mi X, Setoguchi S. Implications of M bias in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2012; 176(10):938–948. [PubMed: 23100247]
16. Pearl J. Invited commentary: understanding bias amplification. *Am J Epidemiol*. 2011; 174(11): 1223–7. [PubMed: 22034488]
17. Patrick AR, Schneeweiss S, Brookhart MA, Glynn RJ, Rothman KJ, Avorn J, Stürmer T. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. *Pharmacoepidemiol Drug Saf*. 2011; 20:551–559. [PubMed: 21394812]
18. Rassen J, Schneeweiss S. Letter to the Editor. *Pharmacoepidemiol Drug Saf*. 2011; 20:1110–1111. [PubMed: 21953847]
19. Toh S, Garcia Rodriguez LA, Hernán MA. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. *Pharmacoepidemiol Drug Saf*. 2011; 20:849–857. [PubMed: 21717528]