On the Use of the Bross Formula for Prioritizing Covariates in the High-Dimensional Propensity Score Algorithm

Richard Wyss[1], Bruce Fireman[2], Jeremy A. Rassen[3], Sebastian Schneeweiss[1]

Author Affiliations:
[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School
[2]Kaiser Permanente, Northern California
[3]Aetion, Inc., New York, New York

In the article titled "*High-dimensional Propensity Score Adjustment in Studies of Treatment Effects Using Health Care Claims Data*," the authors introduce a semi-automated variable selection algorithm for high-dimensional proxy adjustment within insurance healthcare claims databases.[1] The high-dimensional propensity score (HDPS) algorithm evaluates thousands of diagnostic, procedural, and medication claims codes and, for each code, generates binary variables based on the frequency of occurrence for each code during a defined pre-exposure covariate assessment period. The HDPS then prioritizes or ranks each variable based on its potential for bias by assessing the variable's prevalence and univariate association with the treatment and outcome according to the Bross formula.[1, 2] From this ordered list, investigators then specify the number of variables to include in the HDPS model along with pre-specified variables such as age and sex.[1] A full description of the HDPS algorithm is provided elsewhere.[1]

In the original article by Schneeweiss et al.,[1] the Bross bias multiplier for prioritizing covariates was defined as

$$\frac{P_{C1}(RR_{CD}-1)+1}{P_{C0}(RR_{CD}-1)+1}, \text{ if } RR_{CD} \geq 1$$

$$\frac{P_{C1}\left(\frac{1}{RR_{CD}}-1\right)+1}{P_{C0}\left(\frac{1}{RR_{CD}}-1\right)+1}, \text{ if } RR_{CD} < 1,$$

where P_{C1} represents the prevalence of the binary covariate within the exposed group, P_{C0} the prevalence of the binary covariate within the unexposed group, and RR_{CD} the relative risk for the univariate association between the binary covariate and the study outcome.

One of us (BF) noted that for correct assessment of a binary covariate's confounding impact, the Bross bias multiplier should be defined simply as:

$$\frac{P_{C1}(RR_{CD}-1)+1}{P_{C0}(RR_{CD}-1)+1}\text{, for all values of } RR_{CD}$$

We repeated a subset of the analyses from the original manuscript using a revised HDPS that included the correct implementation of the Bross formula. A full description of the data sources that were used for the empirical analyses is provided in the original manuscript.[1, 3] Table 1 shows that there was almost no change from the results reported in the original manuscript after using the above Bross formula for covariate prioritization. For the NSAID data example (Table 1), 199 out of the top 200 ranked variables and 476 out of the top 500 ranked variables were common to both the ordering from the revised HDPS and the ordering from the original manuscript. For the Statin data example (Table 1), 193 out of the top 200 ranked variables and 486 out of the top 500 ranked variables were common to both the ordering from the revised HDPS and the ordering from the original manuscript.

The HDPS software that is distributed online has been updated to include the modified implementation of the Bross formula.[4] Results from analyses that have been conducted using older versions of the HDPS algorithm are unlikely to change meaningfully after this correction.

REFERENCES

1. 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;20:512-522
2. Bross ID. Spurious effects from an extraneous variable. *Journal of chronic diseases*. 1966;19:637-647
3. Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *American journal of epidemiology*. 2011;173:1404-1413
4. Rassen JA, Doherty M, Huang W, Schneeweiss S. Pharmacoepidemiology toolbox. *Boston, MA*.http://www.hdpharmacoepi.org

Table 1. Comparison of results for a subset of analyses reported in Schneeweiss et al. (2009) with results using the revised HDPS

Dataset[a]	Model[b]	No. Covariates Selected[c]	Common HDPS Selected Variables[d]	Results Reported in 2009[1]			Results with Revised HDPS		
				C-Statistic	Odds Ratio	95% CI	C-Statistic	Odds Ratio	95% CI
NSAID									
	1	Unadjusted	-----	------	1.09	0.91, 1.30	------	1.09	0.91, 1.30
	2	$d = 4$	-----	0.61	1.01	0.84, 1.21	0.61	1.01	0.84, 1.21
	3	$d = 4, l = 14$	-----	0.66	0.94	0.78, 1.12	0.66	0.94	0.78, 1.12
	4	$d = 4, l = 14, k = 200$	199 out of 200	0.69	0.86	0.72, 1.04	0.69	0.86	0.72, 1.04
	5	$d = 4, l = 14, k = 500$	476 out of 500	0.71	0.88	0.73, 1.06	0.71	0.87	0.72, 1.06
	5b	$d = 4, k = 500$	476 out of 500	0.71	0.87	0.72, 1.05	0.70	0.88	0.73, 1.06
Statin									
	1	Unadjusted	------	------	0.56	0.51, 0.62	------	0.56	0.51, 0.62
	2	$d = 4$	------	0.70	0.77	0.69, 0.85	0.70	0.77	0.69, 0.85
	3	$d = 4, l = 42$	------	0.82	0.80	0.70, 0.90	0.82	0.80	0.70, 0.90
	4	$d = 4, l = 42, k = 200$	193 out of 200	0.86	0.86	0.76, 0.98	0.85	0.86	0.76, 0.98
	5	$d = 4, l = 42, k = 500$	486 out of 500	0.87	0.86	0.76, 0.98	0.87	0.87	0.76, 0.99
	5b	$d = 4, k = 500$	486 out of 500	0.86	0.89	0.78, 1.02	0.86	0.90	0.79, 1.02

[a] NSAID: comparison of nonsteroidal anti-inflammatory drugs (NSAID) versus selective Cox-2 Inhibitors on GI complications; Statin: comparison of statin use versus glaucoma initiation on risk of death (see Schneeweiss et al. (2009) for further details).

[b] Models 1 through 5b in the above table correspond to Models 1 through 5b in Tables 3 and 4 from the original manuscript by Schneeweiss et al. (2009)

[c] $d =$ the number of demographic variables, $l =$ the number of predefined covariates, and $k =$ the number of empirically selected variables (see Schneeweiss et al. (2009) for further details)

[d] Number of HDPS generated variables that were selected by the revised HDPS (i.e., HDPS with correct implementation of the Bross formula for covariate prioritization) that were also selected by the version of HDPS used in Schneeweiss et al. (2009).