CORRESPONDENCE

Plasma thrombin-cleaved osteopontin as a potential biomarker of acute atherothrombotic ischemic stroke: comments on data sparsity

We were interested to read the paper by Ozaki et al.[1] that was published in *Hypertension Research* in January 2017.[1] The authors aimed to examine the diagnostic performance of the thrombin-cleaved OPN N-terminal (trOPN-N), osteopontin (OPN), metalloproteinase (MMP)-9, S100B, D-dimer and C-reactive protein for the atherothrombotic subtype of ischemic stroke. The results demonstrated that trOPN-N > 5.47 pmol l^{-1} was associated with atherothrombotic risk (odds ratio (OR): 11.7; 95% confidence interval (CI): 2.10–64.87; $P = 0.005$).[1]

The results were interesting; however, it should be noted that the effect of trOPN-N > 5.47 pmol l^{-1} on atherothrombotic risk is biased due to sparse data availability. It is argued that in the presence of data sparsity the OR will be large and its CI tends to be too wide.[2–4] Methods such as *penalization* or *Bayesian* analysis are suggested to remove or limit sparse data bias.[2,3] The authors reported the sensitivity, specificity and *c*-statistic of trOPN-N > 5.47 pmol l^{-1} for atherothrombotic to be 0.54, 0.91 and 0.72, respectively. We drew a 2×2 combination of trOPN-N and atherothrombotic risk using the aforementioned information (Table 1), then we used *penalized* estimation to correct the sparse data bias regarding the effect of trOPN-N > 5.47 pmol l^{-1} on atherothrombotic risk. The corrected OR (95% CI) was found to be 7.73 (2.21, 27.00).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Erfan Ayubi[1,2] and Saeid Safiri[3]

[1]Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [2]Department of Epidemiology & Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran and [3]Managerial Epidemiology Research Center, Department of Public Health, School of Nursing and Midwifery, Maragheh University of Medical Sciences, Maragheh, Iran
E-mail: saeidsafiri@gmail.com

Table 1 The estimated OR from two methods of ordinary logistic regression and penalization

	With athero-thrombotic risk	Without athero-thrombotic risk
trOPN-N > 5.47 pmol l^{-1}	15	3
trOPN-N $\leqslant 5.47$ pmol l^{-1}	13	29
Ordinary logistic regression: estimated OR (95% CI)	11.7 (2.10, 64.87)	
Penalization using data augmentation: estimated OR (95% CI)	10.38 (2.70, 39.87)	

Abbreviations: CI, confidence interval; OR, odds ratio; trOPN-N, thrombin-cleaved OPN N-terminal.

1 Ozaki S, Kurata M, Kumon Y, Matsumoto S, Tagawa M, Watanabe H, Ohue S, Higaki J, Ohnishi T. Plasma thrombin-cleaved osteopontin as a potential biomarker of acute atherothrombotic ischemic stroke. *Hypertens Res* 2017; **40**: 61–66.
2 Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Stat Med* 2015; **34**: 3133–3143.
3 Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ* 2016; **352**: i1981.
4 Ayubi E, Safiri S, Mansournia MA. Increased risk of thromboembolic events in adult congenital heart disease patients with atrial tachyarrhythmias: Bias due to the data sparsity. *Int J Cardiol* 2017; **239**: 20.