Mostofsky et al. (1) published an article in which they presented useful data on the makeup of particle constituents of air pollution, particularly particles with aerodynamic diameters less than 2.5 µm. They also made some claims of an association between various constituents and ischemic stroke. There is some question about whether these claims might be due to multiple testing, as multiple constituents were involved. It is not clear whether ischemic stroke was the only outcome examined because multiple outcomes may have been available but were not reported in their article. Also, there are multiple potential demographic variables that can be used in modeling, so the claims made could be due to model selection bias. Young and Karr (2) discussed potential analysis problems that can occur with observational data sets. Any or all of the 18 chemical constituents could potentially have been used to predict the outcome reported. Any or all of the 11 demographic variables could potentially have been used to adjust the model for confounding variables. The search space for this combination of predictors and adjustors is approximately $(2^{18} - 1) \times (2^{11} - 1) = 5.37 \times 10^8$, or 537 million possible models. It would be useful to others if the authors posted the data set so that multiple modeling issues might be explored.

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES


S. Stanley Young (e-mail: young@niss.org)
Assistant Director of Bioinformatics, National Institute of Statistical Sciences, PO Box 14006, Research Triangle Park, NC 27709

TWO AUTHORS REPLY

We thank Dr. Young for his interest in our study (1); however, his comments (2) are misguided. First, there was no issue of multiple testing because we did not conduct any hypothesis tests and we did not report any P values; we only reported the relative rankings of the β coefficients from different model options. Second, there were no other outcomes—recruitment into the study was based on hospital admission for ischemic stroke. Third, there was no adjustment for demographic variables because this was a self-matched case-crossover study.

Editor’s note: The Journal does not require that authors share their data. Although the sharing of data has many advantages, it is not yet universally regarded by the epidemiologic community as essential for verification. Whether authors should be required to make the raw data available has been discussed elsewhere (Hernán and Wilcox. Epidemiology. 2009;20(2):167–168).

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES


Murray A. Mittleman¹ and Elizabeth Mostofsky²
(e-mail: emostofs@mail.harvard.edu)
¹ Cardiovascular Epidemiology Research Unit, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215
² Department of Epidemiology, Harvard School of Public Health, Boston, MA 02215


DOI: 10.1093/aje/kws460; Advance Access publication: December 27, 2012