Categorizing variables after using regression calibration

Hendriek Boshuizen[1,2] , Stratos-Initiative Topic Group 4 subgroup categorized variables[3]

[1] Biometris/Human Nutrition, Wageningen University and Research, Wageningen, Netherlands

[2] National Institute of Public Health and the Environment, Bilthoven, The Netherlands

[3] Stratos-Initiative Topic Group 4 on measurement error and misclassification, subgroup categorized variables

E-mail for correspondence: hendriek.boshuizen@wur.nl

Abstract:

In nutritional epidemiology consumption of food intake is assessed using instruments (questionnaires, diaries, apps) subject to error. Such measurement error, even when non-differential (i.e. independent of disease status), will bias results of a regression of disease on intake. When intake is used as a continuous measurement, there are many methods available to deliver unbiased estimates, of which regression calibration (replacing the observed intake with a predicted value of true intake, the *calibrated* value) is one of the most popular.

However, in nutritional epidemiology the intake is often categorized, in order to study the shape of the relationship or to avoid assuming linearity. Measurement error in dietary intake will result in misclassification of subjects in these categories. As shown before(1), even when measurement error is non-differential, the misclassification is not.

Here we used a “worst case” simulation to study whether applying categorization on calibrated intake, although theoretically incorrect, would still approximately correct bias in practice, an approach close to that proposed by McMahon (2).

Simulations showed that without confounding this approach works well. However, with strong confounding results are still biased. The likely cause is that a model using categorization mis-specifies the relation between intake and therefore biases the estimated confounder-disease relation.

Key words: regression calibration; measurement error; categorization; variables-with-error; dietary assessment

1) Flegal, K.M., Keyul, P.M. Nieto F.J. Differential Misclassification Arising from Nondifferential Errors in Exposure Measurement. Am J Epidiol. 1991;134(10):1233-44
2) MacMahon, S., Peto, R., Cutler, J., Collins, R., Sorlie, P., Neaton, J., Abbott, R.,Godwin, J., Dyer, A., and Stamler, J. (1990). Blood pressure, stroke, and coronary heart disease. Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet, 335: 765–774.
3) Keogh, R.H., Strawbridge, A,D. , White, I.R. Effects of Classical Exposure Measurement Error on the Shape of Exposure-Disease Associations, Epidemiologic Methods, 2012; 1(1):article 2. DOI: 10.1515/2161-962X.1007