Ann Inst Stat Math (2016) 68:805–825
DOI 10.1007/s10463-015-0511-3

Regression analysis of biased case–control data

Palash Ghosh · Anup Dewanji

Received: 4 September 2013 / Revised: 2 February 2015 / Published online: 17 March 2015

Abstract The data obtained from case–control sampling may suffer from selection or reporting bias, resulting in biased estimation of the parameter(s) of interest by standard analysis of case–control data. In this work, the problem of this bias is dealt with by introducing the concept of reporting probability. Then, considering a reference sample from the source population, we obtain asymptotically unbiased estimate of the population parameters by fitting a pseudo-likelihood, assuming the exposure distribution in the population to be unknown and arbitrary. The proposed estimates of the model parameters follow asymptotically a normal distribution and are semiparametrically fully efficient. We motivate the need for such methodology by considering the analysis of spontaneous adverse drug reaction (ADR) reports in presence of reporting bias.

Keywords Reporting bias · Response-selective sampling · Spontaneous reporting database · Semiparametric estimation · Pseudo-likelihood

P. Ghosh
Centre for Quantitative Medicine, Duke-NUS Graduate Medical School,
Academia-Level 6, 20 College Road, Singapore 169856, Singapore
e-mail: palash.ghosh@duke-nus.edu.sg

A. Dewanji (✉)
Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India
e-mail: dewanjia@isical.ac.in