Check for updates

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Commentary

Statistical Methods in Medical Research
2020, Vol. 29(7) 2034–2035
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0962280219881026
journals.sagepub.com/home/smm
SAGE

Commentary to the recent study by Wang et al.

Toon Braeye[1,2]

In the paper 'Eliminating systematic bias from case-crossover designs', Wang et al.[1] present an issue that is well known in case-crossover research: bias associated with referent selection. In a simulation study, they explore this bias, but because they apply traditional logistic regression instead of conditional logistic regression, their findings contradict previous research.[2,3] Wang et al. state that 'The time-stratified case-crossover design is by no means a final solution to the overlap bias'. Janes et al. state 'As expected, there is no overlap bias for the time stratified and full stratum bidirectional designs'.[4]

Case-crossover referent selection will try to avoid time-varying confounding by the careful selection of referents. A hazard is matched to one or several referents and conditional on these matched sets researches will look into differences between relevant variables associated with the hazard and referent periods. A traditional logistic regression will ignore the matching and will look into differences between all selected referents and all selected hazards.

With time-stratified referent selection (selecting referents on the same weekday, in the same month, same year as the hazard), the number of referents per hazard is not fixed. Depending on the month and year of the hazard, three or four referents are chosen for each hazard. In a simulation scenario in which all days in a period are equally likely hazard days (simulation 1 by Wang et al.), not all days are equally likely selected as referents with a time-stratified referent selection strategy. This will bias coefficients obtained with an 'unconditional' logistic regression analysis. The bias is likely even larger than the bias observed after ambidirectional selection of referents (a referent selection strategy known to result in overlap bias), because with this referent selection strategy, an equal number of referents, four, is selected for each hazard, and the difference between the distributions of relevant covariates associated with referents and hazards is only caused by observations at the start and end of the period under investigation.

In conclusion, time-stratified referent selection strategies are free from overlap bias when analyzed with conditional logistic regression, but a bias will present itself after analysis with traditional logistic regression. To illustrate this point, we added a conditional logistic regression to the simulation study of Wang et al. (Figure 1, upper row). We feel it would be confusing to also name the bias from applying traditional instead of conditional logistic regression 'overlap' bias.

In addition, the solution to this bias suggested by the authors, the 'calibration approach', is only correct when the 'true' effect under investigation equals zero. To illustrate this, we have added a scenario to simulation 1, in which the 'true' effect for O_3 equals 0.1 (instead of 0 as in the simulation study by Wang et al.). The calibration approach did only slightly improve coefficient estimation but did not remove the bias. Conditional logistic regression allowed for an accurate estimation of the O_3-effect (Figure 1, lower row). (As the different exposures (CO, NO, NO_2, etc.) are correlated and the models always only include only one of these variables (as in the simulation study by Wang), there is always some bias in coefficient estimation for the exposures that are not O_3 in the lower row of Figure 1.)

(The R code of this simulation study will be uploaded as part of the submission.)

[1]Department Epidemiology of infectious diseases, Epidemiology, Sciensano, Brussels, Belgium
[2]Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Hasselt, Belgium

Corresponding author:
Toon Braeye, Department Epidemiology of infectious diseases, Epidemiology, Sciensano; Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Juliette Wytsmanstraat 14, Brussels 1050, Belgium.
Email: toon.braeye@sciensano.be

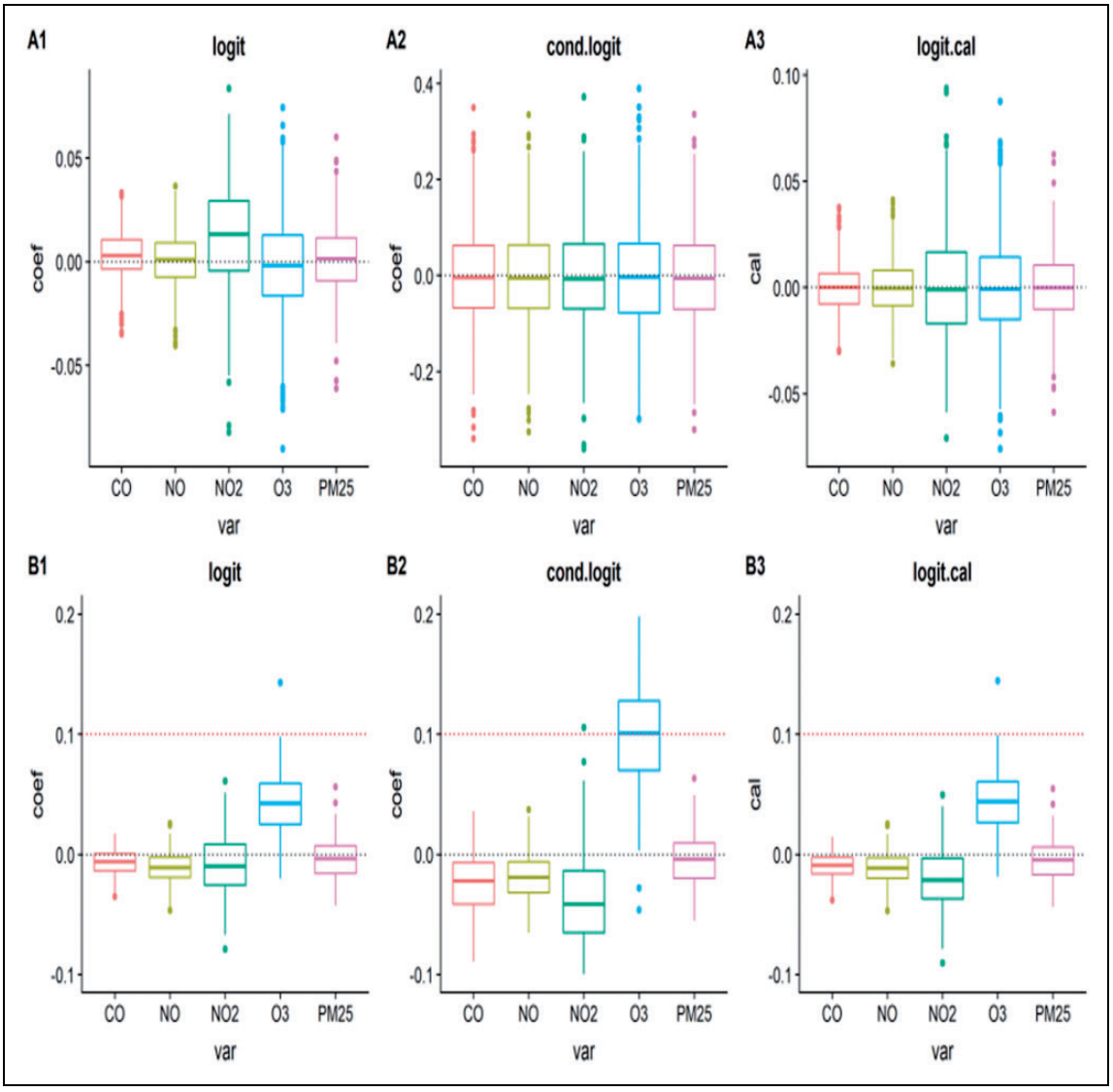

Figure 1. All coefficients = 0 (a). The coefficient for $O_3 = 0.1$ all other coefficients = 0 (b). Traditional logistic regression (logit, 1), conditional logistic regression (cond.logit, 2) and logistic regression in combination with calibration (logit.cal, 3).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.