Bias and Estimation under Misspecification of the Risk Period in Self-Controlled Case Series Studies

Supplemental Information

Luis Fernando Campos, Damla Şentürk, Yanjun Chen, and Danh V. Nguyen

Appendix

Proof of the bias function (3). The relative incidence estimator (2) is

$$\widehat{R}^* = \frac{\sum_{i=1}^N \tilde{n}_{i1} / \sum_{i=1}^N \tilde{n}_{i0}}{\tilde{e}_1/\tilde{e}_0} = \frac{n_{..}^{-1} \sum_{i=1}^N \tilde{n}_{i1} / \{1 - n_{..}^{-1} \sum_{i=1}^N \tilde{n}_{i1}\}}{(\tau + u)/(e_0 - u)}$$

where $n_{..} = \sum_{ik} n_{ik}$. By the law of large numbers and Slutsky's theorem, $\widehat{R}^*$ is consistent for $R^* = \{\tilde{\pi}_1/(1-\tilde{\pi}_1)\}/\{(\tau+u)/(e_0-u)\}$. Direct calculation yields $\tilde{\pi}_1 = (u + \tau R)/\Delta$, where $\Delta = e_0 + \tau R$; thus, $R^* = (u + \tau R)/\tilde{\tau}$, which a linear function of the inverse of the specified risk length, $1/\tilde{\tau}$. Since $u = \tilde{\tau} - \tau$, we can express R^* as

$$R^* = R + \tau(R-1)\left(\tilde{\tau}^{-1} - \tau^{-1}\right) \equiv \gamma_0 + \gamma_1\left(\tilde{\tau}^{-1} - \tau^{-1}\right), \qquad \text{for } \tilde{\tau} > \tau$$

where $\gamma_0 = R$ and $\gamma_1 = \tau(R-1)$.

Next, consider the case when $u < 0$ (i.e., $\tilde{\tau} < \tau$). Note that $e_0 = T - \tau$, where T is the total follow-up time and also $u = \tilde{\tau} - \tau$. Similar to the case with $u > 0$, application of the the law of large numbers and Slutsky's theorem, gives that $\widehat{R}^*$ is consistent for $R^* = \{\tilde{\pi}_1/(1-\tilde{\pi}_1)\}/\{\tilde{\tau}/(T-\tilde{\tau})\}$, where $\tilde{\pi}_1 = (\tau+u)R/\{(T-\tau) - uR + \tau R\}$ and we have expressed $\tilde{e}_1/\tilde{e}_0 = \tilde{\tau}/(T-\tilde{\tau})$. Simplification gives $R^* = (e_0 - u)R/(e_0 - uR)$ and we can express this as a nonlinear function of $1/\tilde{\tau}$ as follows:

$$\begin{aligned}
R^* &= R + \frac{(e_0-u)R}{e_0 - uR} - R = R + \frac{uR(R-1)}{e_0 - uR} = \frac{(R-1) - \tau(R-1)\tilde{\tau}^{-1}}{\frac{T+\tau(R-1)}{R}\tilde{\tau}^{-1} - 1} \\
&= R + \frac{(R-1) - \tau(R-1)\tau^{-1} - \tau(R-1)(\tilde{\tau}^{-1} - \tau^{-1})}{\frac{T+\tau(R-1)}{R}(\tilde{\tau}^{-1} - \tau^{-1}) + \frac{T+\tau(R-1)}{R}\tau^{-1} - 1} \\
&= R + \frac{-\tau(R-1)(\tilde{\tau}^{-1} - \tau^{-1})}{\frac{T+\tau(R-1)}{R}(\tilde{\tau}^{-1} - \tau^{-1}) + \frac{T-\tau}{\tau R}} \equiv \gamma_0 + \frac{-\gamma_1\left(\tilde{\tau}^{-1} - \tau^{-1}\right)}{\gamma_2(\tilde{\tau}^{-1} - \tau^{-1}) + \gamma_3},
\end{aligned}$$

where $\gamma_2 = \{T + \tau(R-1)\}/R$ and $\gamma_3 = (T-\tau)/(\tau R)$. Thus, combining the above expressions for R^* for both cases of $u > 0$ and $u < 0$, one obtains equation (3) in Section (2.1):

$$R^* = \gamma_0 + \gamma_1 \left(\tilde{\tau}^{-1} - \tau^{-1}\right)_- + \frac{-\gamma_1 \left(\tilde{\tau}^{-1} - \tau^{-1}\right)_+}{\gamma_2(\tilde{\tau}^{-1} - \tau^{-1})_+ + \gamma_3},$$

where $(x)_+ = x$ if $x > 0$ and 0 otherwise and, similarly, $(x)_- = x$ if $x < 0$ and 0 otherwise. It is clear that $R^* = \gamma_0 = R$ when $\tilde{\tau}^{-1} = \tau^{-1}$ (i.e., $u = 0$). This result for the functional form of the bias due to risk length misspecification motivated our proposed estimation procedure.

Jacobian matrix for equation (8). For the Jacobian matrix $\mathbf{J}$ in (8), define $h_l = h_l(\beta^*, \boldsymbol{\alpha}^*)$, for $l = 0, \ldots, J$, $\pi^*_{ij\cdot} = \sum_{k=1}^{1} \pi^*_{ijk}$ and $\pi^*_{i\cdot k} = \sum_{j=1}^{J} \pi^*_{ijk}$. With these simplifications, direct calculations give the m^{th} row of $\mathbf{J}$ as $(\partial h_m/\partial\beta^*, \partial h_m/\partial\alpha^*_1, \ldots, \partial h_m/\partial\alpha^*_J)$, where $\partial h_0/\partial\beta^* = -\sum_{i=1}^{N} n_{i\cdot\cdot}\pi^*_{i\cdot 1}(1 - \pi^*_{i\cdot 1})$, $\partial h_0/\partial\alpha^*_j = \sum_{i=1}^{N} n_{i\cdot\cdot}(\pi^*_{ij1} - \pi^*_{ij\cdot}\pi^*_{i\cdot 1})$ for $j = 1, \ldots, J$, $\partial h_j/\partial\beta^* = -\sum_{i=1}^{N} n_{i\cdot\cdot}(\pi^*_{ij1} - \pi^*_{ij\cdot}\pi^*_{i\cdot 1})$, $\partial h_j/\partial\alpha^*_j = -\sum_{i=1}^{N} n_{i\cdot\cdot}\pi^*_{ij\cdot}\ (1 - \pi^*_{ij\cdot})$, for $j = 1, \ldots, J$, $\partial h_j/\partial\alpha^*_l = \sum_{i=1}^{N} n_{i\cdot\cdot}\pi^*_{ij\cdot}\pi^*_{il\cdot}$ for $j \neq l$. The above expressions are similar to the ones given in Mohammed et al. (2013).

Reference

Mohammed, SM, Dalrymple, DS, Senturk, D & Nguyen, DV (2013) 'Naive hypothesis testing for case series models with time-varying exposure onset measurement error: Inference for infection-cardiovascular risk in patients on dialysis', *Biometrics*, **69**, 520–529.

Table S1: Summary of the 32 simulation experiments and study design parameters (i)-(iv). For each experiment, the true relative incidence, $R = \exp(\beta)$, varies from 0.7 to 4.

Parameter	Values
(i) Number of exposures	(a) Single exposure, similar to MMR-ITP data or (b) Multiple exposures: 1, 2 or 3 with probabilities 0.7, 0.2 and 0.1, respectively
(ii) Number of age groups, J	No age group; 2 age groups; 3 age groups ($J = 0, 1, 2$)
Age group relative incidences, $\exp(\alpha_j)$	$(e^{\alpha_0}, e^{\alpha_1}) = (1,\ 1.35)$, $(e^{\alpha_0}, e^{\alpha_1}, e^{\alpha_2}) = (1,\ 1.35,\ 1.65)$
(iii) Length of risk period, τ	15, 30, 45 days
(iv) Distribution of exposure times	Uniform, Normal
Risk period relative incidences, $\exp(\beta)$	0.7, 0.9, 1.2, 1.5, 2, 4
Sample size, N	100, 200, 400, 800 individuals