Robust Cox Regression as an Alternative Method to Estimate Adjusted Relative Risk in Prospective Studies with Common Outcomes

Wuxiang Xie[1,*] and Fanfan Zheng[2,*]

[1]*Department of Epidemiology and Biostatistics, Imperial College London, UK*

[2]*Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China*

Abstract: *Objective*: To demonstrate the use of robust Cox regression in estimating adjusted relative risks (and confidence intervals) when all participants with an identical follow-up time and when a common outcome is investigated.

Methods: In this paper, we propose an alternative statistical method, robust Cox regression, to estimate adjusted relative risks in prospective studies. We use simulated cohort data to examine the suitability of robust Cox regression.

Results: Robust Cox regression provides estimates that are equivalent to those of modified Poisson regression: regression coefficients, relative risks, 95% confidence intervals, *P* values. It also yields reasonable probabilities (bounded by 0 and 1). Unlike modified Poisson regression, robust Cox regression allows for four automatic variable selection methods, it directly computes adjusted relative risks for continuous variables, and is able to incorporate time-dependent covariates.

Conclusion: Given the popularity of Cox regression in the medical and epidemiological literature, we believe that robust Cox regression may gain wider acceptance and application in the future. We recommend robust Cox regression as an alternative analytical tool to modified Poisson regression. In this study we demonstrated its utility to estimate adjusted relative risks for common outcomes in prospective studies with two or three waves of data collection (spaced similarly).

Keywords: Robust Cox regression, Modified Poisson regression, Logistic regression, Relative risk, Odds ratio.

BACKGROUND

The relative risk is commonly used in prospective studies to determine the effect of an exposure on a binary outcome. When confounding exists, the unadjusted relative risk cannot correctly estimate the effect of the exposure, and an adjusted relative risk should be estimated. When researchers use data from two waves of collection to evaluate the effect of an exposure on an outcome, logistic regression is generously used. For instance, Johnsen *et al.* (2005) used a cohort study to explore whether monocyte count is a predictor of novel plaque formation among 2610 persons without carotid plaque at baseline [1]. The baseline ultrasound examination was conducted in 1994-1995, with a follow-up in 2000-2001 [1]. In this article, logistic regression was used to calculate an odds ratio, but the ratio may overestimate the relative risk because the rate of plaque formation (40%) is a common outcome [1, 2].

Several other methods also can be used to estimate adjusted relative risks, such as the Mantel–Haenszel method, log-binomial regression, and Poisson regression. However, the Mantel–Haenszel method cannot adjust for continuous covariates; log-binomial regression does not always converge well; [3] and Poisson regression provides a wider 95% confidence interval [3]. To solve these problems, modified (or robust) Poisson regression with robust error variance was proposed and first published in 2004 by Zou [3]. Zou adopted the SAS PROC GENMOD procedure with the REPEATED statement to obtain unbiased estimates of adjusted relative risks [3]. Modified Poisson regression converges more reliably than log-binomial regression, and displays more accurate 95% confidence intervals than Poisson regression [3]. Since its publication, modified Poisson regression has gained popularity in medical and public health research. By September 30, 2016, Zou's paper had been cited by more than 2500 scientific publications. However, modified Poisson regression has a number of practical limitations. First, the method permits probability estimates exceeding 1, which is implausible. Second, the SAS PROC GENMOD procedure does not allow for automatic variable selection methods (such as stepwise selection), which are sometimes desirable. Third, the PROC GENMOD procedure provides regression coefficients but cannot directly provide adjusted relative risk estimates for continuous explanatory variables—an extra antilogarithm transformation is needed to obtain these. Furthermore,

*Address correspondence to these authors at the Department of Epidemiology and Biostatistics, School of Public Health, Medical School, St Mary's Campus, Imperial College London, UK; Tel: 020 7594 3328;
Fax: 020 7594 3456; E-mail: w.xie@imperial.ac.uk

Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China; Tel: 86 010 82544774; Fax: 86 010 82544777;
E-mail: fanfan.zheng@nlpr.ia.ac.cn

E-ISSN: 1929-6029/16

when repeated data from three or more waves are used to evaluate the effect of an exposure, time-dependent covariates may exist. However, modified Poisson regression is not able to incorporate time-dependent covariates.

When time-to-event survival data are used, Cox regression is generally regarded as an appropriate approach to estimate the effects of an exposure. In time-to-event studies, subjects are followed-up for the outcomes of interest. Survival (or failure) times are censored when subjects are lost to follow-up, when participants die for reasons other than the outcome of interest, or when the study is completed before the outcome is reached. In prospective studies, when all participants are assigned the same follow-up time, the hazard ratio equals the relative risk [4, 5]. Although the Cox model yields correct point estimates, it tends to overestimate the variance in coefficients, leading to wider confidence intervals when follow-up time is the same [6]. This problem can be solved by computing robust variance estimates, which was initially proposed to obtain a robust inference for Cox regression when using small sample sizes [7]. Barros and Hirakata (2003) have demonstrated that robust Cox regression is equivalent to modified Poisson regression when analyzing binary outcomes in cross-sectional studies—the parameters estimated are the same [6]. Although Barros and Hirakata recommended that robust Cox regression could also be used in longitudinal studies when a constant risk period is assigned to all participants, this method was not well accepted in the applied statistical community. It has also seldom been used in medical and public health research. Therefore, we conducted this study to demonstrate the use of robust Cox regression in estimating adjusted relative risks (and confidence intervals) in cases with identical follow-up times and in the presence of time-dependent covariates. The regression was implemented using the SAS PROC PHREG procedure with the COVSANDWICH statement. We compared robust Cox regression with other available methods based on both simulated and real data sets.

METHODS

The Partial Likelihood of Robust Cox Regression

The Cox regression model is formulated as

$$\lambda_i(t) = \lambda(t; Z_i) = \lambda_0(t)\, exp(\beta Z_i),$$

where $\lambda_0(t)$ is an arbitrary and unspecified baseline hazard function (the hazard function when all covariates equal 0), Z_i is a vector of explanatory variables for the ith individual, and β is a vector of regression parameters associated with the explanatory variables [8]. The partial likelihood function is formulated as

$$PL = \prod_{i=1}^{n} [\frac{\exp(\beta Z_i)}{\sum_{j \in R(y_i)} \exp(\beta Z_j)}]^{\delta_i},$$

where δ_i is an indicator variable with a value of 1 if t_i is uncensored or a value of 0 if t_i is censored, and $j \in R(y_i)$ is a convenient mechanism for excluding from the denominator participants who already experienced the event.

When a constant risk period is assigned to all participants, the partial likelihood will be reduced to a very simple form. This is achieved by using the Breslow method to handle the tie, as follows,

$$PL = \frac{\exp(\beta S)}{[\sum_i \exp(\beta Z_i)]^m},$$

where m is the number of the outcome of interest, and $S = Z_1 + Z_2 + \ldots + Z_m$.

The Advantages of Robust Cox Regression over Modified Poisson Regression in Practice

First, the probabilities estimated by robust Cox regression never exceed 1, because Cox regression estimates individual probabilities by subtracting the survival function from 1. More specifically,

$$Probability = 1 - S(t; Z_i) = 1 - S_0(t)^{\exp(\beta Z_i)},$$

where $S_0(t)$ is the baseline survival function at observation time t. Clearly, the probability is bounded by 0 and 1 because the survival rate (calculated by the survival function) must lie within this range. Second, the PROC PHREG procedure allows choice between four automatic variable selection methods: forward selection, backward elimination, stepwise selection, and best subsets selection. The SELECTION code in the MODEL statement is used to enable this. Finally, the robust Cox models directly compute adjusted relative risks for continuous and categorical variables [9].

RESULTS

Simulation Study 1: Using Data from Two Surveys to Evaluate the Effect of an Exposure

For illustrative purpose, we created several hypothetical studies, each focusing on the association

Table 1: Results from Simulation Study 1: Adjusted Relative Risks (RR), Odds Ratios (OR) and 95% Confidence Intervals (CI) for Studies of Common Outcomes

True RR	Stratum-specific risk				Methods											
	C=1		C=0		Logistic		Mantel–Haenszel		Log-binomial		Modified Poisson		Cox*		Robust Cox*	
	E=1	E=0	E=1	E=0	OR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
0.33	0.2	0.6	0.1	0.3	0.19	0.12, 0.31	0.33	0.24, 0.46	0.33	0.24, 0.46	0.33	0.24, 0.46	0.33	0.23, 0.48	0.33	0.24, 0.46
0.50	0.2	0.4	0.1	0.2	0.40	0.25, 0.63	0.50	0.35, 0.71	0.50	0.35, 0.71	0.50	0.35, 0.71	0.50	0.34, 0.74	0.50	0.35, 0.71
1.00	0.4	0.4	0.2	0.2	1.00	0.67, 1.49	1.00	0.77, 1.31	1.00	0.77, 1.30	1.00	0.77, 1.31	1.00	0.72, 1.39	1.00	0.77, 1.31
2.00	0.8	0.4	0.4	0.2	4.05	2.72, 6.03	2.00	1.61, 2.48	2.00	1.62, 2.48	2.00	1.61, 2.49	2.00	1.50, 2.66	2.00	1.61, 2.49
3.00	0.6	0.2	0.3	0.1	5.00	3.20, 7.81	3.00	2.15, 4.19	3.00	2.15, 4.19	3.00	2.15, 4.19	3.00	2.05, 4.39	3.00	2.15, 4.19

Abbreviations: C, confounder; E, exposure.
*Assigning a constant time to all participants.

Table 2: Results from Simulation Study 1: Six Methods to Evaluate Risk Bias and 95% Confidence Interval (CI) Coverage Based on 1,000 Runs

True RR	Stratum-specific risk				Methods											
	C=1		C=0		Logistic		Mantel–Haenszel		Log-binomial		Modified Poisson		Cox*		Robust Cox*	
	E=1	E=0	E=1	E=0	relative bias%	95% CI converge	relative bias%	95% CI converge	relative bias%	95% CI converge	relative bias%	95% CI converge	relative bias%	95% CI converge	relative bias%	95% CI converge
0.33	0.2	0.6	0.1	0.3	–42.45	33.5	–0.38	95.1	–0.08	94.5	–0.36	95.2	–0.36	97.9	–0.36	95.2
0.50	0.2	0.4	0.1	0.2	–20.40	82.6	0.00	94.8	0.14	94.8	–0.15	94.7	–0.15	97.2	–0.15	94.7
1.00	0.4	0.4	0.2	0.2	0.00	94.3	0.00	94.5	–0.44	94.0	0.00	94.4	0.00	97.9	0.00	94.4
2.00	0.8	0.4	0.4	0.2	102.63	5.0	0.14	95.8	–0.08	95.1	0.07	95.5	0.07	98.7	0.07	95.5
3.00	0.6	0.2	0.3	0.1	69.01	37.0	0.71	96.4	0.70	96.5	0.90	96.8	0.90	98.3	0.90	96.8

Abbreviations: C, confounder; E, exposure.
*Assigning a constant time to all participants.

between exposure to a specific risk factor and disease, and each requiring adjustment for a confounder. The sample size of 500 was assigned a relative risk of 0.33, 0.50, 1.00, 2.00, or 3.00. The subjects were equally split into two groups—one with a confounder, the other without (not a dummy variable). Of the 250 subjects in the group with a confounder, 150 were assigned to the exposure group. In the group without a confounder, 100 subjects were assigned to the exposure group.

The stratum-specific risks, calculated adjusted odds ratios, relative risks, and 95% confidence intervals derived from each method are shown in Table **1** (logistic regression, the Mantel–Haenszel test, log-binomial regression, modified Poisson regression, Cox regression, and robust Cox regression). When the true relative risk was 1, the adjusted odds ratio estimated by logistic regression equaled the relative risk (Table **1**). However, the odds ratio overestimated the true relative risk when it exceeded 1, and underestimated the true relative risk when it was below 1 (Table **1**). All the other methods yielded the same point relative risks (Table **1**). Although Cox regression estimated wider 95% confidence intervals, the confidence intervals estimated by robust Cox regression were equivalent to those obtained using modified Poisson regression (Table **1**).

We used a macro (Macro 1 in the Appendix) to create 1,000 random data sets according to stratum-specific risks. The sample size for each run was set at 500. For each statistical method, the adjusted relative risks were estimated on the 1,000 random data sets. Here, the relative bias percentage was calculated as follows: [(median of adjusted relative risk estimated from 1,000 random data sets − true relative risk) / true relative risk] × 100. The coverage of 95% confidence interval is defined as the proportion of runs with an estimated 95% confidence interval including the true relative risk. The simulation results are shown in Table **2**. We note that the risk bias of robust Cox regression was equivalent to that of modified Poisson regression. As expected**,** the Cox regression produces wider confidence intervals for the adjusted relative risk. However, the results demonstrate that the coverage percentage obtained by robust Cox regression was close to 95%, and also equivalent to that of modified Poisson regression (Table **2**).

Simulation Study 2: Using Data from Three Surveys with the Same Time Interval to Evaluate the Effect of an Exposure

When data are collected in three waves (spaced similarly), two different time values exist. For example, we conducted a cohort study to evaluate the effect of hypertension on the incidence of carotid plaque. The first survey was conducted in 2002, with follow-ups in 2007 and 2012. Participants were included in analyses if they were without plaques at baseline and attended at least two surveys. There are therefore five categories of participants and two follow-up times, as presented in Table **3**.

In this simulation study, the 500 simulated subjects according to the structure in **Simulation study 1** were assigned to the group with 10 years of follow-up. Another group containing 250 subjects with 5 years of follow-up was also assigned a relative risk of 0.33, 0.50, 1.00, 2.00, or 3.00. The 250 subjects were equally split into two groups: one with a confounder, the other without. Of the 125 subjects in the group with a confounder, 50 were assigned to the exposure group. In the group without a confounder, 75 subjects were assigned to the exposure group. We used a macro (Macro 2 in the Appendix) to create 1,000 random data sets according to stratum-specific risk (Table **4**).

As shown in Table **4**, after adjusting for the confounder and follow-up time, the point relative risks and 95% confidence intervals estimated by robust Cox regression (assigning a constant time to all participants) were equivalent to those estimated by modified Poisson regression. When using the follow-up time as the failure time variable, robust Cox regression estimated relative risks with relatively larger relative bias percentages.

Simulation Study 3: Using Data from Three Surveys with the Same Time Interval to Evaluate the Effect of an Exposure when a Time-Dependent Covariate Exists

Time-dependent covariates are those that may change in value over the follow-up period. For example, some participants without hypertension at baseline may become hypertensive during the first 5-year follow-up (see Table **3**). Therefore, these participants should be analyzed as part of the exposure group during the second 5-year follow-up. Based on **Simulation study 2**, we used a macro (Macro 3 in the Appendix) to create 1,000 random data sets. These assumed that 80% of participants with 10-year follow-ups were in the 5th category, and 30% of participants without hypertension at baseline became hypertensive during the first 5-years of follow-up. As expected, the median adjusted relative risk was closer to 1 (Table **5**), because the participants that were reclassified in the

Table 3: Five Categories of Participants Identified in a Prospective Cohort Study to Evaluate the Effect of Hypertension on Carotid Plaque Formation

Categories	Risk factor and carotid ultrasound survey			Follow-up time (years)
	In 2002	In 2007	In 2012	
1st	Attended	Attended	Non-attended	5
2nd	Non-attended	Attended	Attended	5
3rd	Attended	Non-attended	Attended	10
4th	Attended	Attended and detected a new plaque	Attended	5
5th	Attended	Attended and did not detect a new plaque	Attended	10

Table 4: Results from Simulation Study 2: Modified Poisson Regression and Robust Cox Regression to Evaluate Risk Bias and 95% Confidence Interval (CI) Coverage, Based on 1,000 Runs

True RR	Stratum-specific risk								Methods					
	Follow-up time=10				Follow-up time=5				Modified Poisson*		Robust Cox**		Robust Cox†	
	C=1		C=0		C=1		C=0							
	E=1	E=0	E=1	E=0	E=1	E=0	E=1	E=0	relative bias%	95% CI converge	relative bias%	95% CI converge	relative bias%	95% CI converge
0.33	0.2	0.6	0.1	0.3	0.1	0.3	0.05	0.15	0.13	95.3	0.13	95.3	–4.18	95.4
0.50	0.2	0.4	0.1	0.2	0.1	0.2	0.05	0.1	0.75	94.4	0.75	94.4	–3.49	93.7
1.00	0.4	0.4	0.2	0.2	0.2	0.2	0.1	0.1	–0.06	94.8	–0.06	94.8	–3.65	94.1
2.00	0.8	0.4	0.4	0.2	0.4	0.2	0.2	0.1	0.17	94.0	0.17	94.0	–3.11	93.7
3.00	0.6	0.2	0.3	0.1	0.3	0.1	0.15	0.05	0.58	96.0	0.58	96.0	–2.08	95.2

Abbreviations: C, confounder; E, exposure.
*Adjusting the follow-up time as a confounder.
**Assigning a constant time to all participants, and adjusting the follow-up time as a confounder.
†Using the follow-up time as the failure time variable.

Table 5: Results from Simulation Study 3: Using Robust Cox Regression to Evaluate Relative Risk (RR) when a Time-Dependent Covariate Exists (1,000 Runs)

Assigned RR	Stratum-specific risk								Median of adjusted RR
	Follow-up time=10				Follow-up time=5				
	C=1		C=0		C=1		C=0		
	E=1	E=0	E=1	E=0	E=1	E=0	E=1	E=0	
0.33	0.2	0.6	0.1	0.3	0.1	0.3	0.05	0.15	0.41
0.50	0.2	0.4	0.1	0.2	0.1	0.2	0.05	0.1	0.55
1.00	0.4	0.4	0.2	0.2	0.2	0.2	0.1	0.1	0.95
2.00	0.8	0.4	0.4	0.2	0.4	0.2	0.2	0.1	1.78
3.00	0.6	0.2	0.3	0.1	0.3	0.1	0.15	0.05	2.60

Abbreviations: C, confounder; E, exposure.

second 5-year follow-up had a higher risk when assigned a relative risk <1, or a lower risk when assigned a relative risk >1. The SAS code used to estimate adjusted relative risks when a time-dependent variable exists is presented in the Appendix.

DISCUSSION

In this paper, we proposed robust Cox regression as a convenient and efficient relative risk estimation approach for prospective studies of common outcomes. This was demonstrated using simulated and real

examples where two or three waves of data were used (with similar time intervals between them). Cox regression is widely used in cohort studies and clinical trials; we suggest that robust Cox regression could also be used in these contexts, and would be readily understood.

Logistic regression is widely used to study the factors associated with binominal outcomes, adjusting for covariates. When the incidence is low (<10%), the odds ratio is close to the relative risk. However, for common outcomes (>10%), the odds ratio always overestimates the relative risk (when it is more than 1) or underestimates the relative risk (when it is less than 1) [2]. To solve this problem, Zhang and Yu (1998) proposed a method that corrects the adjusted odds ratio from logistic regression [2]. They also derived an estimate of an association or treatment effect that better represents the true relative risk [2]. Nevertheless, McNutt *et al.* (2003) revealed a potential bias in their method, and proposed that unbiased estimates of relative risk in common outcome studies could be obtained from log-binomial and Poisson regression analyses [10]. However, both of these alternatives are known to be unsatisfactory. Log-binominal regression is prone to convergence problems, while Poisson regression generates wider confidence intervals [3]. To help resolve this, Zou proposed modified Poisson regression with robust error variance, which has been widely accepted and applied in recent years [3]. Subsequent studies have proposed modified log-binominal regression using the COPY method [11], nonlinear programming [12], and Bayesian methods [13] to estimate the relative risk in prospective studies. However, these three methods have yet to be widely used in medical and public health research. The first method is inherently limited by PROC GENMOD and yields estimators based on an expanded dataset rather than the original dataset [11]. The main disadvantage of the latter two methods is their lack of familiarity and accessibility to researchers.

Cox regression is widely used for survival analysis with censored data. Lee (1994) was the first to recommend Cox regression for estimating prevalence ratios in cross-sectional studies, but wider standard errors were obtained [4]. Barros and Hirakata tried to use robust Cox regression to estimate prevalence ratios in cross-sectional studies, and found that the estimates were the same as those obtained from modified Poisson regression [6], which our results support. Additionally, we identified four advantages of robust Cox regression over modified Poisson regression. First, the method restricts the estimated probabilities to a realistic range. Second, four automatic variable selection methods can be used, and are readily available in standard statistical packages. Third, the method directly provides adjusted relative risk estimates for continuous explanatory variables. Finally, it has the capability to incorporate time-dependent covariates.

We have used the robust Cox regression in the analysis of several cohort studies. In all cases we used the same modeling strategy and generated the same results of relative risk. We consider that robust Cox regression could be a useful and alternative method to help identify abnormal results from robust Poisson regression when assessing the predicted individual probabilities. In this article, we proposed to apply robust Cox model to analysis binary data. Meanwhile, the robust error estimate is widely used to deal with variance underestimation in correlated data analysis. Future studies should be focus on applying the method with correlated binary data from cohort studies or clinical trials. On the other hand, the virtue of the Cox regression is its capability to incorporate time-dependent covariates. It is thus interesting to investigate whether robust Cox model can still retain the interpretation of relative risk with time-dependent covariates.

CONCLUSION

In summary, estimates provided by robust Cox regression are equivalent to those obtained from modified Poisson regression, while the former also confers several advantages. Given the popularity of Cox regression in epidemiologic research, we believe that robust Cox regression may become widely used and accepted in the near future. We therefore recommend robust Cox regression as an alternative approach to estimate adjusted relative risks in prospective studies. Findings from this study particularly support its use when analyzing data from two or three waves, when the same interval is used between waves, and when a common outcome is used.

COMPETING INTERESTS

None declared.

FUNDING

This study was funded by the Newton International Fellowship (project no. P56804) and National Natural Science Foundation of China (project no. 81302503 and 81601176).

APPENDIX

Macro 1 for Simulation Study 1

```
%let seed=12345;
%macro ds (risk1, risk2, risk3, risk4);
%do k =1 %to 1000;
data ds&k;
  do id =1 to 500;
    if id<=150 then do; confound=1;expose=1;disease=ranbin(&seed,1,&risk1);end;
          else if 150<id<=250 then do; confound=1;expose=0;disease=ranbin(&seed,1,&risk2);end;
          else if 250<id<=350 then do; confound=0;expose=1;disease=ranbin(&seed,1,&risk3);end;
          else if id>350 then do; confound=0;expose=0;disease=ranbin(&seed,1,&risk4);end;
    runs=&k;
          output;
  end;
run;
%let seed=%eval(12345+&k);
%end;
%mend;
%ds(risk1=0.2, risk2=0.6, risk3=0.1, risk4=0.3);
```

Macro 2 for simulation study 2

```
%let seed=12345;
%macro ds (risk1, risk2, risk3, risk4, risk5, risk6, risk7, risk8);
%do k =1 %to 1000;
data ds&k;
  do id =1 to 750;
    if id<=150 then do;time=10;confound=1;expose=1;disease=ranbin(&seed,1,&risk1);end;
  else if 150<id<=250 then do;time=10;confound=1;expose=0;disease=ranbin(&seed,1,&risk2);end;
    else if 250<id<=350 then do;time=10;confound=0;expose=1;disease=ranbin(&seed,1,&risk3);end;
    else if 350<id<=500 then do;time=10;confound=0;expose=0;disease=ranbin(&seed,1,&risk4);end;
          else if 500<id<=550 then do;time=5;confound=1;expose=1;disease=ranbin(&seed,1,&risk5);end;
          else if 550<id<=625 then do;time=5;confound=1;expose=0;disease=ranbin(&seed,1,&risk6);end;
          else if 625<id<=700 then do;time=5;confound=0;expose=1;disease=ranbin(&seed,1,&risk7);end;
          else if id>700 then do;time=5;confound=0;expose=0;disease=ranbin(&seed,1,&risk8);end;
    runs=&k;
          output;
  end;
run;
%let seed=%eval(12345+&k);
%end;
%mend;
%ds(risk1=0.2, risk2=0.6, risk3=0.1, risk4=0.3, risk5=0.1, risk6=0.3, risk7=0.05, risk8=0.15);
```

Macro 3 for simulation study 3

```
%let seed=12345;
%macro ds (risk1,risk2,risk3,risk4,risk5,risk6,risk7,risk8);
%do k =1 %to 1000;
data ds&k;
 do id =1 to 750;
    if id<=150 then do;time=10;confound=1;expose=1;disease=ranbin(&seed,1,&risk1);end;
    else if 150<id<=250 then do;time=10;confound=1;expose=0;disease=ranbin(&seed,1,&risk2);end;
    else if 250<id<=350 then do;time=10;confound=0;expose=1;disease=ranbin(&seed,1,&risk3);end;
    else if 350<id<=500 then do;time=10;confound=0;expose=0;disease=ranbin(&seed,1,&risk4);end;
    else if 500<id<=550 then do;time=5;confound=1;expose=1;disease=ranbin(&seed,1,&risk5);end;
    else if 550<id<=625 then do;time=5;confound=1;expose=0;disease=ranbin(&seed,1,&risk6);end;
    else if 625<id<=700 then do;time=5;confound=0;expose=1;disease=ranbin(&seed,1,&risk7);end;
    else if id>700 then do;time=5;confound=0;expose=0;disease=ranbin(&seed,1,&risk8);end;
    runs=&k;
 x=ranbin(&seed,1,0.8);
```

```
 if ((x=1 and time=10) and expose=0) then x2=ranbin(&seed,1,0.3);else x2=0;
     output;
   end;
run;
data ds&k;
   set ds&k;
 expose2=expose;
 start=0;
 if x2=0 then do;
     disease2=disease;
     stop=time;
          output;
   end;
 else do;
 stop=5;
 disease2=0;
 output;
 start=5;
 stop=10;
 expose2=1;
      disease2=disease;
      output;
   end;
run;
%let seed=%eval(12345+&k);
%end;
%mend;
%ds(risk1=0.2, risk2=0.6, risk3=0.1, risk4=0.3, risk5=0.1, risk6=0.3, risk7=0.05, risk8=0.15);
```

SAS code for estimating adjusted RR when a time-dependent variable exists

```
ods output ParameterEstimates= robustcox;
proc phreg data= database covsandwich;
 class expose2(reference='0');
 model (start,stop)*disease2(0)=expose2 confound/rl ties=BRESLOW;
 by runs;
data robustcox;
 set robustcox;
 if Parameter="confound" then delete;
proc means data= robustcox median;
 var HazardRatio;
run;
```

REFERENCES

Johnsen SH, Fosse E, Joakimsen O, Mathiesen EB, Stensland-Bugge E, Njølstad I, Arnesen E. Monocyte count is a predictor of novel plaque formation: a 7-year follow-up study of 2610 persons without carotid plaque at baseline the Tromso Study. Stroke 2005; 36: 715-9. https://doi.org/10.1161/01.STR.0000158909.07634.83

Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA 1998; 280: 1690-1. https://doi.org/10.1001/jama.280.19.1690

Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004; 159: 702-6. https://doi.org/10.1093/aje/kwh090

Lee J. Odds ratio or relative risk for cross-sectional data? Int J Epidemiol 1994; 23: 201-3. https://doi.org/10.1093/ije/23.1.201

Breslow N. Covariance analysis of censored survival data. Biometrics 1974; 30: 89-99. https://doi.org/10.2307/2529620

Barros AJ, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Med Res Methodol 2003; 3: 21. https://doi.org/10.1186/1471-2288-3-21

Lin DY, Wei LJ. The Robust Inference for the Cox Proportional Hazards Model. Journal of the American Statistical Association 1989; 84: 1074-8. https://doi.org/10.1080/01621459.1989.10478874

Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society 1972; 34: 187-220.

SAS Institute I: SAS/STAT sofeware, version 9.2. Cary, NC: SAS Institute, Inc 2008.

McNutt LA, Wu C, Xue X, *et al*. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol 2003; 57: 940-3. https://doi.org/10.1093/aje/kwg074

Deddens JA, Petersen MR. Approaches for estimating prevalence ratios. Occup Environ Med 2008; 65: 481-6. https://doi.org/10.1136/oem.2007.034777

Yu B, Wang Z. Estimating relative risks for common outcome using PROC NLP. Comput Methods Programs Biomed 2008; 90: 179-86. https://doi.org/10.1016/j.cmpb.2007.12.010

Chu H, Cole SR. Estimation of risk ratios in cohort studies with common outcomes: a Bayesian approach. Epidemiology 2010; 21: 855-62. https://doi.org/10.1097/EDE.0b013e3181f2012b

Received on 25-10-2016 Accepted on 28-11-2016 Published on 09-12-2016

http://dx.doi.org/10.6000/1929-6029.2016.05.04.1