

## Fitting Parametric and Semi-parametric Conditional Poisson Regression Models with Cox's Partial Likelihood in Self-controlled Case Series and Matched Cohort Studies

Stanley Xu<sup>1</sup>, Paul Gargiullo<sup>2</sup>, John Mullooly<sup>3</sup>, David McClure<sup>1</sup>,  
Simon J. Hambidge<sup>1,4,5</sup> and Jason Glanz<sup>1</sup>

<sup>1</sup>*Kaiser Permanente Colorado*, <sup>2</sup>*CDC*, <sup>3</sup>*Kaiser Permanente Northwest*, <sup>4</sup>*University of Colorado* and <sup>5</sup>*Denver Health*

*Abstract:* The self-controlled case series (SCCS) and the matched cohort are two frequently used study designs to adjust for known and unknown confounding effects in epidemiological studies. Count data arising from these two designs may not be independent. While conditional Poisson regression models have been used to take into account the dependence of such data, these models have not been available in some standard statistical software packages (e.g., SAS). This article demonstrates 1) the relationship of the likelihood function and parameter estimation between the conditional Poisson regression models and Cox's proportional hazard models in SCCS and matched cohort studies; 2) that it is possible to fit conditional Poisson regression models with procedures (e.g., *PHREG* in SAS) using Cox's partial likelihood model. We tested both conditional Poisson likelihood and Cox's partial likelihood models on data from studies using either SCCS or a matched cohort design. For the SCCS study, we fitted both parametric and semi-parametric models to model age effects, and described a simple way to apply the parametric and complex semi-parametric analysis to case series data.

*Key words:* Conditional Poisson regression, Cox's partial likelihood, matched cohort studies, self-controlled case series, semi-parametric model.

### 1. Introduction

Epidemiological and medical studies frequently use count data in which exposed and unexposed incidence rates are compared. The data are usually analyzed with Poisson regression models, and the observations may not be independent due to clustering naturally or by study design. For example, in studies of vaccine safety, a method known as the self-controlled case series (SCCS) has been used. The SCCS is a case-only method in which a subject's follow-up period is

partitioned into exposed and unexposed intervals. Incidence rates in risk periods following vaccination are compared to those in unexposed control periods outside of the risk periods, and each case acts as its own control (Farrington, 1995; Farrington, Nash and Miller, 1996; Kramarz *et al.*, 2000; France *et al.*, 2004; Glanz *et al.*, 2006). SCCS data are analyzed with conditional Poisson regression models to account for the within-subject dependence. Conditioning is used to obtain a likelihood in which only cases need to be sampled. As a result, intercept parameters explaining each subject's individual baseline risk for the event count of interest is not present in the likelihood (Farrington, 1995; Roy, *et al.*, 2006). This results in a de facto adjustment for all subject-level risk factors and potential confounders (measured and not measured), allowing only within-subject comparisons of incidence rates between exposed and unexposed time intervals. This is similar to a stratified analysis with each subject as a unique stratum. SCCS is particularly useful for controlling for confounding by indication, wherein the probability of exposure to vaccine is related to the subject-level risk of the outcome (Whitaker, *et al.*, 2006).

Another example of dependent count data is the matched cohort method, in which exposed subjects are matched to unexposed subjects on certain characteristics to reduce confounding effects. Matched cohort methods have been applied to injury studies (Walker *et al.*, 1981; Walker *et al.*, 1983; Kannus, *et al.*, 2000) and to studies of traffic crashes (Evans, 1986; Cummings, McKnight and Weiss 2003; Cummings, McKnight and Greenland, 2003) where drivers and passengers are naturally matched on the vehicle-related variables such as vehicle model, speed, place, time, and type of crash. The matched subjects form a unit called a cluster or stratum. The events within the strata are often considered dependent, and the data are also analyzed with conditional Poisson regression models (Cameron and Trivedi, 1998, chap. 9; Diggle, 2002, chap. 5).

In addition to conditional Poisson regression models, proportional hazards models using Cox's partial likelihood can be used to analyze matched count data. It is well known that conditional logistic regression can be performed using a stratified Cox's proportional hazards model, resulting in a product partial likelihood with strata representing matched sets (Cox, 1975; Hosmer and Lemeshow, 1999). Similarly, by conditioning on the total number of events in each stratum, the conditional Poisson likelihood function is of a product multinomial likelihood (Agresti, 2002, chap. 8). It is claimed that the stratified Cox's partial likelihood with an arbitrary constant as the time to event gives the same results as a conditional Poisson regression model (Cummings, McKnight and Weiss 2003; Cummings, McKnight and Greenland, 2003). The Cox's stratified partial likelihood has been used in vaccine safety studies for modeling count data (France *et al.*, 2004; Glanz *et al.*, 2006; Hambidge, *et al.*, 2006). But this has not been

mathematically proven and studied in depth. Our study examines this claim and shows why conditional Poisson regression models and Cox's stratified proportional hazard models produce the same coefficient estimates and variances in SCCS and matched cohort studies.

We first demonstrate the relationship between the conditional Poisson likelihood and Cox's stratified partial likelihood for count data in the SCCS study design. We will then extend this comparison to the matched cohort study design. We will then give examples of each study design, and finish with a discussion of implications of our findings.

## 2. SCCS Study

### 2.1 Conditional Poisson likelihood and Cox's stratified partial likelihood

Suppose that the  $i$ th subject is followed for a period of time,  $t_i$  days, in which there are  $t_{ie}$  exposure days and  $t_{iu}$  unexposed days,  $t_i = t_{ie} + t_{iu}$ . Let  $z_{ij}$  denote an indicator variable equal to 1 for an event and 0 for no event at day  $j$  for subject  $i$ . Then  $y_{ie} = \sum_{j=1}^{t_{ie}} z_{ij}$  is the number of events in the exposed period and  $y_{iu} = \sum_{j=1}^{t_{iu}} z_{ij}$  is the number of events in the unexposed period.

According to Farrington (1995), the conditional Poisson likelihood (CL) is the product of the likelihood across subjects which is of the following form for the  $i$ th subject

$$L_{CL(i)} = \left\{ \frac{t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta})}{t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta}) + t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})} \right\}^{y_{ie}} \left\{ \frac{t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})}{t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta}) + t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})} \right\}^{y_{iu}} \quad (2.1)$$

where  $x_{ie}$  is a row vector of covariates for the exposed period and  $x_{iu}$  is a row vector of covariates for the unexposed period. In equation (2.1) we have defined only two time periods: exposed and unexposed. Even if the exposure status remains unchanged within a time interval, further partitioning of the interval would be necessary for changes in values of time-varying covariates other than exposure status. In other words, a covariate can be time-varying across the exposed and unexposed periods, but constant within each period of time. Otherwise, more partitions are needed for fixed covariate values with an interval. If a covariate is constant over both exposed and unexposed periods, it will cancel out of equation (2.1).

If each day of follow-up is treated as an observation for subject  $i$ , and a dummy variable (e.g., *surv*) for the time to event in survival analyses is assigned a value of 1 (or any positive constant) for each observation, all of the days in the follow-up period are at risk. Further, unlike conventional survival analysis,

assume that each individual  $i$  is a member of all  $t_i$  risk sets. That is, there is no dropping out of risk sets for any reason including becoming an event or being censored, until the end of the observation period. Within a subject-level stratum, the Cox's (Cox, 1975) partial likelihood (PL) is then

$$\begin{aligned} L_{PL(i)} &= \prod_{j=1}^{t_i} \left\{ \frac{\exp(\mathbf{x}_{ij}\boldsymbol{\beta})}{\sum_{j=1}^{t_i} \exp(\mathbf{x}_{ij}\boldsymbol{\beta})} \right\}^{z_{ij}} = \prod_{j=1}^{t_{ie}} \left\{ \frac{\exp(\mathbf{x}_{ij}\boldsymbol{\beta})}{\sum_{j=1}^{t_i} \exp(\mathbf{x}_{ij}\boldsymbol{\beta})} \right\}^{z_{ij}} \prod_{j=1}^{t_{iu}} \left\{ \frac{\exp(\mathbf{x}_{ij}\boldsymbol{\beta})}{\sum_{j=1}^{t_i} \exp(\mathbf{x}_{ij}\boldsymbol{\beta})} \right\}^{z_{ij}} \\ &= \left\{ \frac{\exp(\mathbf{x}_{ie}\boldsymbol{\beta})}{t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta}) + t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})} \right\}^{y_{ie}} \left\{ \frac{\exp(\mathbf{x}_{iu}\boldsymbol{\beta})}{t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta}) + t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})} \right\}^{y_{iu}} \quad (2.2) \end{aligned}$$

Note that equations (2.1) and (2.2) differ in the numerators of the two terms. Equation (2.2) has no person times in the numerators while equation (2.1) does.

## 2.2 Maximum likelihood estimates (MLEs)

The conditional Poisson likelihood and Cox's partial likelihood are used in place of likelihood for obtaining estimates for  $\boldsymbol{\beta}$  and their standard errors. Taking logs of equations (2.1) and (2.2),  $\log L_{CL(i)} = \log L_{PL(i)} + y_{ie} \log(t_{ie}) + y_{iu} \log(t_{iu})$ , it becomes obvious that the first and second derivatives with respect to  $\boldsymbol{\beta}$  are identical. It can be shown that the first and second derivatives of  $\log L_{CL(i)}$  and  $\log L_{PL(i)}$  are

$$\begin{aligned} \frac{\partial \log L_{(i)}}{\partial \boldsymbol{\beta}} &= \frac{y_{ie} \mathbf{x}_{ie} t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta}) - y_{ie} \mathbf{x}_{iu} t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta})}{t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta}) + t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})} \\ &\quad + \frac{y_{iu} \mathbf{x}_{iu} t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta}) - y_{iu} \mathbf{x}_{ie} t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})}{t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta}) + t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})} \\ \frac{\partial^2 \log L_{(i)}}{\partial \boldsymbol{\beta}^2} &= \frac{t_{iu} t_{ie} \exp(2\mathbf{x}_{iu}\boldsymbol{\beta}) (2y_{ie} \mathbf{x}'_{iu} \mathbf{x}_{ie} + 2y_{iu} \mathbf{x}'_{iu} \mathbf{x}_{ie} - y_{ie} \mathbf{x}'_{ie} \mathbf{x}_{ie})}{\{t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta}) + t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})\}^2} \\ &\quad + \frac{t_{iu} t_{ie} \exp(2\mathbf{x}_{ie}\boldsymbol{\beta}) (-y_{ie} \mathbf{x}'_{iu} \mathbf{x}_{iu} - y_{iu} \mathbf{x}'_{ie} \mathbf{x}_{ie} - y_{iu} \mathbf{x}'_{iu} \mathbf{x}_{iu})}{\{t_{ie} \exp(\mathbf{x}_{ie}\boldsymbol{\beta}) + t_{iu} \exp(\mathbf{x}_{iu}\boldsymbol{\beta})\}^2} \end{aligned}$$

When there is only an exposure indicator variable in the covariate vector (e.g.,  $x_{ie} = 1$  and  $x_{iu} = 0$ ), the first and second derivatives reduce to

$$\frac{\partial \log L_{(i)}}{\partial \boldsymbol{\beta}} \Big|_{x_{ie}=1, x_{iu}=0} = \frac{y_{ie} t_{iu} - y_{iu} t_{ie} \exp(\boldsymbol{\beta})}{t_{ie} \exp(\boldsymbol{\beta}) + t_{iu}}$$

and

$$\frac{\partial^2 \log L_{(i)}}{\partial \boldsymbol{\beta}^2} \Big|_{x_{ie}=1, x_{iu}=0} = -\frac{t_{iu} t_{ie} \exp(\boldsymbol{\beta}) (y_{ie} + y_{iu})}{\{t_{ie} \exp(\boldsymbol{\beta}) + t_{iu}\}^2}$$

To obtain  $\hat{\beta}$ ,

$$\begin{aligned} \frac{\partial \log L}{\partial \beta} &= \sum_{i=1}^n \frac{y_{ie} \mathbf{x}_{ie} t_{iu} \exp(\mathbf{x}_{iu} \beta) - y_{ie} \mathbf{x}_{iu} t_{ie} \exp(\mathbf{x}_{iu} \beta)}{t_{ie} \exp(\mathbf{x}_{ie} \beta) + t_{iu} \exp(\mathbf{x}_{iu} \beta)} \\ &+ \sum_{i=1}^n \frac{y_{iu} \mathbf{x}_{iu} t_{ie} \exp(\mathbf{x}_{ie} \beta) - y_{iu} \mathbf{x}_{ie} t_{ie} \exp(\mathbf{x}_{ie} \beta)}{t_{ie} \exp(\mathbf{x}_{ie} \beta) + t_{iu} \exp(\mathbf{x}_{iu} \beta)} = 0 \end{aligned}$$

where  $n$  is the number of subjects. Since there is no closed form of solution for  $\beta$ , either nonlinear optimization program or EM algorithm can be used to obtain the maximum likelihood (partial) estimate of  $\beta$ . The standard error of the estimate of  $\beta$  is the square root of this variance,

$$Var(\hat{\beta}) = \text{inv} \left\{ -E \left( \frac{\partial^2 \log L}{\partial \beta^2} \right) \right\}_{\beta=\hat{\beta}}$$

More generally, the variance-covariance matrix for a vector of parameter estimates for a multivariate model is the inverse of the negative symmetric matrix of partial second derivatives with respect to all pairs of parameters (Hessian matrix).

The difference in the likelihood function has no impact on fixed effect estimates and their standard errors. If all of the subjects have the same exposed and unexposed person-times ( $t_e$  and  $t_u$  for all subjects) and there is only an exposure indicator in the covariate vector, then the effect of exposure and its variance are

$$\hat{\beta} = \log \frac{\sum_{i=1}^n y_{ie} t_u}{\sum_{i=1}^n y_{iu} t_e}$$

and

$$Var(\hat{\beta}) = \frac{1}{\sum_{i=1}^n \frac{t_{iu} t_{ie} \exp(\hat{\beta})(y_{ie} + y_{iu})}{\{t_{ie} \exp(\hat{\beta}) + t_{iu}\}^2}} = \frac{\{t_e \exp(\hat{\beta}) + t_u\}^2}{\sum_{i=1}^n t_u t_e \exp(\hat{\beta})(y_{ie} + y_{iu})}$$

### 3. Matched Cohort Study

The purpose of a matched cohort study is to control confounding effects by matching exposed to unexposed subjects on certain characteristics. Usually a common correlation is assumed among the subjects within a stratum. Let  $k$  denote the  $k$ th stratum, in which there are  $n_k$  subjects who have the same values of matching criteria. Similar to the SCCS study design, the conditional Poisson likelihood for the  $k$ th stratum in a matched cohort study is then

$$L_{CL(k)} = \prod_{i=1}^{n_k} \frac{t_{ki} \exp(\mathbf{x}_{ki}) \beta}{\prod_{i=1}^{n_k} t_{ki} \exp(\mathbf{x}_{ki}) \beta} \tag{3.1}$$

where  $y_{ki}$  is the dependent variable and  $t_{ki}$  is the duration for subject  $i$  in the  $k$ th stratum. If the unit of  $t$  is the day and an event can only occur once in a day, we can treat each day as an observation, and then there are  $t_{ki}$  observations for the  $i$ th subject in the  $k$ th stratum. The Cox's stratified partial likelihood will be

$$L_{PL(k)} = \prod_{i=1}^{n_k} \frac{\exp(\mathbf{x}_{ki})\boldsymbol{\beta}}{\prod_{i=1}^{n_k} t_{ki} \exp(\mathbf{x}_{ki})\boldsymbol{\beta}} \quad (3.2)$$

We showed in section 2.2, that the difference between (3.1) and (3.2) has no impact on the estimation of fixed effects. Thus, we expect the same estimates of fixed effects and their standard errors from the conditional Poisson regression and the Cox's stratified partial likelihood method in matched cohort studies.

## 4. Examples

### 4.1 A SCCS study

Miller *et al.* (2001) studied the association between measles, mumps, rubella (MMR) vaccine and idiopathic thrombocytopenic purpura (ITP) in children aged 12-23 months. The data were later updated in a SCCS tutorial by Whitaker *et al.* (2006) The ITP events are defined as hospital admissions for ITP. Six of the 35 children were admitted to hospital more than once during the follow up period. In the tutorial, six age groups were used for a parametric model: 366-426 days, 427-487 days, 488-548 days, 549-609 days, 610-670 days, and 671-730 days. However, mis-specification of the age groups can produce biased estimation of the association of ITP with MMR. A semi-parametric model was proposed in which the age effect was left unspecified (Whitaker, *et al.*, 2006; Farrington and Whitaker, 2006). In our current analysis, we fitted both parametric and semi-parametric models in analyzing the MMR-ITP data. Three risk windows were studied: 0-14, 15-28, and 29-42 days after vaccination. The days outside of the risk windows represent the control window. A sample patient's MMR-ITP data were used to demonstrate how the data were expanded. This patient was followed from day 366 to 730, was vaccinated at day 710, and experienced two events at days 414 and 418. Computer programs (e.g., SAS) can be used to produce the expanded data with three risk windows, five indicator variables for age groups (671-730 day is the reference age group), and 43 indicator variables for the days that ITP events occurred.

Fitting the newly developed semi-parametric model is computationally challenging (Whitaker, *et al.*, 2006). It is shown in Table 1 that both conditional Poisson regression and Cox's stratified partial likelihood method give the same level of association and 95% confidence intervals between MMR and ITP for

the three risk periods after MMR vaccination for both parametric and semi-parametric models although the  $-2 \log$  likelihood ( $-2\text{ll}$ ) values differ. Coefficients for age categories in the parametric models are the same for both methods (data not shown). The parametric model (subject-stratified) for the expanded MMR-ITP data included the following indicator variables risk1, risk2, risk3, age1, age2, age3, age4 and age5. For the semi-parametric model (subject-stratified), indicator variables Iday1-Iday43 were added to covariates. Coefficients for all 43 events days were given in the above model. In the semi-parametric model from STATA, 42 coefficients were given relative to the one of Iday1 (set to zero). The same values of 42 coefficients relative to Iday1 can be obtained by subtracting a coefficient with the one of Iday1. For example the coefficients for Iday1 and Iday2 are 19.377 and 19.287, respectively. The coefficient of Iday2 relative to Iday1 is  $19.287-19.377 = -0.09$ , which is identical to the output from STATA. Statistical test (e.g., *test* statement in PHREG in SAS) can be used to examine if the relative coefficient equals to zero.

Table 1: Relative incidences (95% CI) for analyses of ITP and MMR\*

Methods		Risk period: days after MMR			-2 ll**
		0-14	15-28	29-42	
PL <sup>+</sup>	parametric	1.31 (0.30, 5.73)	5.95 (2.52, 14.07)	2.60 (0.75, 9.07)	486.7
	semi-parm	1.46 (0.32, 6.60)	5.45 (2.17, 13.66)	2.04 (0.56, 7.50)	361.2
CL <sup>++</sup>	parametric	1.31 (0.30, 5.73)	5.95 (2.52, 14.07)	2.60 (0.75, 9.07)	226.8
	semi-parm	1.46 (0.32, 6.60)	5.45 (2.17, 13.66)	2.04 (0.56, 7.50)	299.7

\*, days outside of the risk windows represent the control window.

\*\* ,  $-2 \log PL$  or  $-2 \log CL$ .

<sup>+</sup>, Cox's stratified partial likelihood method fitted in SAS for both parametric and semi-parametric models.

<sup>++</sup>, conditional Poisson likelihood regression fitted in SAS for the parametric model and in STATA for the semi-parametric model.

## 4.2 A matched cohort study

This study compares health services use and cost between obese and non-obese managed care organization (MCO) members (Raebel, *et al.*, 2004). A total of 539 obese and 1,225 non-obese members were matched by age, sex, and medical clinic with matching ratios from 1 to 3. One outcome of interest is professional service claims over a one year period. Letting  $y_{ki}$  be the number of professional service claims over a one year period ( $t_{ki}$ ) for member  $i$  in the  $k$ th stratum, we assumed that the member had events (professional service claims) in  $y_{ki}$  days, and no events in  $(365 - y_{ki})$  days. The source data does not identify on which day professional service claims occurred and they were arbitrarily assigned to the days during the follow-up period. The ordering of days is arbitrary and irrelevant to

the Cox's stratified partial likelihood method. The results from the conditional Poisson likelihood regression and Cox's stratified partial likelihood method are shown in Table 2.

Table 2: Relative incidences (95% CI) comparing professional services claims between obese and non-obese HMO members.

Methods	Relative incidences	-2 ll*
PL <sup>+</sup>	1.79 (1.57, 2.04)	13197.0
CL <sup>++</sup>	1.79 (1.57, 2.04)	2223.2

\*,  $-2 \log PL$  or  $-2 \log CL$ .

<sup>+</sup>, Cox's stratified partial likelihood method fitted in SAS.

<sup>++</sup>, conditional Poisson likelihood regression fitted in SAS.

## 5. Discussion

To our knowledge, this is the first paper to compare the likelihood of a conditional Poisson regression model to partial likelihood of a stratified proportional hazards regression model in both a matched cohort and SCCS analysis. Despite the differences in the likelihood functions, the estimates and standard errors of coefficients from the conditional Poisson regression models and the stratified proportional hazards models are equivalent. If time intervals in the conditional Poisson regression models are expressed as single days, the likelihoods become identical. Theoretically, we proved that the maximum likelihood estimates from these two approaches are the same because the first and second derivatives of the log-likelihood function, with respect to the coefficients, are identical. Using actual MMR vaccination and ITP case information, we also published a simulation study (Glanz, *et al.*, 2006) in which data were simulated based on different risk levels (relative incidence=1.5, 2.0, 3.0, 4.0), with varying person times across the subjects, and with and without time-varying covariates (seasonality). The simulated data sets contain no 'real' health care information. The simulation results showed that the two approaches generated identical results under all of the conditions.

Conditional Poisson regression is a useful method for estimating relative incidences in a matched pair setting such as SCCS and matched cohort studies. More recently, Roy *et al.* (2006) developed new estimation strategies that allow endogenous time-varying covariates and missing at random dropouts in conditional Poisson models. Establishing conditions under which conditional Poisson regression models may be implemented using a Cox's proportional hazard framework would be beneficial for several reasons. First, unlike procedures for Cox's proportional hazard models, conditional Poisson regression models are not universally

available in commercial software packages. For instance, conditional Poisson regression models are offered in STATA (StataCorp, 2003) but not in SAS (SAS, 2002). Although conditional Poisson models are coded in a series of macros in SAS from the website <http://statistics.open.ac.uk/sccs>, it requires the knowledge of SAS macros and running a series of SAS macros. It has also been shown that maximizing the log conditional Poisson likelihood is equivalent to maximizing the Poisson log-likelihood with a fixed intercept entered in the model for each subject in a SCCS study or for each cluster in a matched cohort study (Lancaster, 2002). However, the recently developed semi-parametric models, which are designed to avoid estimation bias from the mis-specification of age groups, pose computational challenge and can not be easily fitted in SAS. It is shown in the MMR-ITP example that the complex semi-parametric models can be easily fitted using the SAS procedure PHREG. Regardless of the software package, most statisticians are quite familiar with statistical procedures for fitting Cox's proportional hazard models since it is the most commonly used method for performing time-to-event analyses. In particular, statistical analysts in the field of vaccine safety could benefit from understanding the relationship between the conditional Poisson regression models and the stratified Cox's proportional hazard models and, furthermore, how to create the datasets and use Cox's partial likelihood to obtain relative incidence estimates in SCCS studies while adjusting other time-varying covariates such as seasonality and ages.

### Acknowledgements

This work was funded through a subcontract with America's Health Insurance Plans (AHIP) under contract 200-2002-00732 from the Centers for Disease Control and Prevention (CDC).

### References

- Agresti, A. (2002). *Categorical Data Analysis*, second edition. Wiley.
- Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*, Econometric Society Monograph No.30. Cambridge University Press.
- Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 269-76.
- Cummings, P., McKnight, B. and Greenland, S. (2003). Matched cohort methods for injury research. *Epidemiologic Review* **25**, 43-50.
- Cummings, P., McKnight, B. and Weiss, N. S. (2003). Matched-pair cohort methods in traffic crash research. *Accident Analysis and Prevention* **35**, 131-41.
- Diggle, P. J., Heagerty, P., Liang, K-Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford University Press,.

- Evans, L. (1986). The effectiveness of safety belts in preventing fatalities. *Accident Analysis and Prevention* **18**, 229-41.
- Farrington, C. P. (1995). Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* **51**, 228-35.
- Farrington, C. P., Nash, J. and Miller, E. (1996). Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am Journal of Epidemiology* **143**, 1165-73.
- Farrington, C. P. and Whitaker, H. J. (2006). Semiparametric analysis of case series data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **55**, 553-594.
- France, E. K., Glanz, J. M., Xu, S., Davis, R. L., Black, S. B., Shinefield, H. R., angwill, K. M., Marcy, S. M., Mullooly, J. P., Jackson, L. A. and Chen, R. (2004 ). Safety of the trivalent inactivated influenza vaccine among children: A population-based study. *Archives of Pediatrics and Adolescent Medicine* **158**, 1031-36.
- Glanz, J. M., McClure, D. L., Xu, S., Hambidge, S. J., Lee, M., Kolczak, M. S., Kleinman, K., Mullooly, J. P. and France, E. K. (2006). Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations. *Journal of Clinical Epidemiology* **59**, 808-818.
- Hambidge, S. J., Glanz, J. M., France, E. K., McClure, D., Xu, S., Yamasaki, K., Jackson, L., Mullooly, J. P., Zangwill, K.M., Marcy, S.M., Black, S. B., Lewis, E. M., Shinefield, H. R., Belongia, E., Nordin, J., Chen, R. T., Shay, D. K., Davis, R. L., DeStefano, F. and the Vaccine Safety Datalink Team. (2006). Safety of trivalent inactivated influenza vaccine in children 6 to 23 months old. *Journal of the American Medical Association* **296**(16), 1990-1997.
- Hosmer, D. W. Jr, Lemeshow, S. (1999). *Applied survival analysis: regression modeling of time to event data*. John Wiley.
- Kannus, P., Parkkari, J., Niemi, S., Pasanen, M., Palvanen, M., Jarvinen, M. and Vuori, I. (2000). Prevention of hip fracture in elderly people with use of a hip protector. *New England Journal of Medicine* **343**, 1506-13.
- Kramarz, P., DeStefano, F., Gargiullo, P. M., Davis, R. L., Chen, R. T., Mullooly, J. P., Black, S. B., Shinefield, H. R., Bohlke, K., Ward, J. I. and Marcy, M.S. (2000). Does influenza vaccination exacerbate asthma? Analysis of a large cohort of children with asthma. Vaccine Safety Datalink Team. *Archives of Family Medicine* **9**, 617-23.
- Lancaster, T. (2002). Orthogonal Parameters and Panel Data. *Review of Economic Studies* **69**, 647-666.
- Miller, E., Waight, P., Farrington, P., Stowe, J. and Taylor, B. (2001). Idiopathic thrombocytopenic purpura and MMR vaccine. *RArchives of Disease in Childhood* **84**, 227-229.

- 
- Raebel, M. A., Malone, D. C., Conner, D. A., Xu, S., Porter, J. A. and Lanty, F. A. (2004). Health services use and health care costs of obese and non-obese individuals. *Archives of Internal Medicine* **164**, 2135-40.
- Roy, J., Alderson, D., Hogan, J. W. and Tashima, K. T. (2006). Conditional Inference Methods for Incomplete Poisson Data With Endogenous Time-Varying Covariates: Emergency Department Use Among HIV-Infected Women. *Journal of the American Statistical Association* **101**, 424-434
- SAS 9.1. (2002). SAS Institute Inc. Cary, NC, USA.
- StataCorp. Stata (2003 ). Statistical Software: Release 8.0. College Station, TX: Stata Corporation.
- Walker, A. M., Jick, H., Hunter, J. R., Danford, A., Watkins, R. N., Alhadeff, L. and Rothman, K. J. (1981). Vasectomy and non-fatal myocardial infarction. *Lancet* **1**, 13-15.
- Walker, A. M., Jick, H., Hunter, J. R. and McEvoy, J 3rd. (1983). Vasectomy and nonfatal myocardial infarction: continued observation indicates no elevation of risk. *Journal of Urology* **130**, 936-7.
- Whitaker, H. J., Farrington, C. P., Spiessens, B. and Musonda, P. (2006). Tutorial in Biostatistics: The self-controlled case series method. *Statistics in Medicine* **25**, 1768-1797.

Received June 3, 2008; accepted November 8, 2008.

Stanley Xu  
Kaiser Permanente Colorado  
P.O. Box 378066  
Denver, Colorado 80237, U.S.A.  
Stan.Xu@kp.org

Paul Gargiullo  
Vaccine Safety DataLink Team  
Immunization Safety Office  
CDC, Atlanta, Georgia, USA  
PGargiullo@cdc.gov

John Mullooly  
Center for Health Research  
Kaiser Permanente Northwest  
Portland, Oregon, USA  
John.Mullooly@kpchr.org

David McClure  
Kaiser Permanente Colorado  
P.O. Box 378066  
Denver, Colorado 80237, U.S.A.  
David.L.McClure@kp.org

Simon J. Hambidge  
Department of Pediatrics  
University of Colorado School of Medicine  
Denver, Colorado, USA  
Simon.J.Hambidge@kp.org

Jason Glanz  
Kaiser Permanente Colorado  
P.O. Box 378066  
Denver, Colorado 80237, USA  
Jason.M.Glanz@kp.org