Lifetime Data Anal (2007) 13:333–350
DOI 10.1007/s10985-007-9045-1

Cox regression analysis in presence of collinearity: an application to assessment of health risks associated with occupational radiation exposure

Xiaonan Xue · Mimi Y. Kim · Roy E. Shore

Received: 3 August 2006 / Accepted: 27 June 2007 / Published online: 28 July 2007

Abstract This paper considers the analysis of time to event data in the presence of collinearity between covariates. In linear and logistic regression models, the ridge regression estimator has been applied as an alternative to the maximum likelihood estimator in the presence of collinearity. The advantage of the ridge regression estimator over the usual maximum likelihood estimator is that the former often has a smaller total mean square error and is thus more precise. In this paper, we generalized this approach for addressing collinearity to the Cox proportional hazards model. Simulation studies were conducted to evaluate the performance of the ridge regression estimator. Our approach was motivated by an occupational radiation study conducted at Oak Ridge National Laboratory to evaluate health risks associated with occupational radiation exposure in which the exposure tends to be correlated with possible confounders such as years of exposure and attained age. We applied the proposed methods to this study to evaluate the association of radiation exposure with all-cause mortality.

Keywords Ridge regression · Collinearity · Cox proportional hazards model · Occupational exposure

X. Xue (✉) · M. Y. Kim
Division of Biostatistics, Department of Epidemiology & Population Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Belfer 1303C, Bronx, 10461, NY, USA
e-mail: xxue@aecom.yu.edu

M. Y. Kim
e-mail: mikim@aecom.yu.edu

R. E. Shore
Radiation Effects Research Foundation, 5-2 Hijiyama Koen, Minami-ku, Hiroshima, 732-0815, Japan
e-mail: shore@rerf.or.jp

1 Introduction

In epidemiologic studies, collinearity among covariates in the model is quite common. Additional factors besides exposure to the putative hazardous agent are frequently considered in occupational risk assessment models, such as attained age, calendar year, sex, job category, age at exposure, length of employment, initial employment year and years since termination of employment (Wing and Richardson 1999; Gilbert et al. 1993a, b; Cardis et al. 1995). Some of these variables, such as length of employment, length of time since initial employment and attained age, tend to be correlated with each other and may also be correlated with cumulative radiation exposure. Collinearity can induce unstable estimates of covariate effects and may speciously mask or amplify the putative radiation effect in the analysis (Marsh et al. 2002). However, exclusion of such variables may result in a biased risk estimator due to residual "healthy worker survivor" effects (i.e., healthy workers tend to work a longer time and thereby get more exposure). The collinearity problem has seldom been considered in occupational radiation studies. Gilbert (1989) pointed out that controlling for variables that correlate strongly with cumulative occupational exposure often results in loss of power; however, no correction methods were provided. Categorizing the variables that are correlated with each other may reduce the correlation, but often results in a nontrivial loss of efficiency (Lagakos 1988).

Ridge regression has been used as an alternative estimation method in multiple linear regression models when multicollinearity exists among covariates. With multicollinearity, a ridge type estimator is suggested because it has a smaller total mean square error (MSE) than the maximum likelihood estimator. When the multicollinearity is large, the reduction in MSE can be sizeable. The ridge regression estimator was first proposed by Hoerl and Kennard (1970a, b) for linear regression models and was later generalized to logistic regression (Schafer et al. 1984). Barker and Brown (2001) performed a sequence of simulations to evaluate ridge estimators for logistic regression analysis when binary predictor variables are highly correlated. The ridge regression method has been applied to a study on correlated nutrients and colon cancer (Smith et al. 1991) and a study of joint effects of correlated PCB congeners on breast cancer risk (Holford et al. 2000). However, this method has not been adapted for Cox regression models (Cox 1972).

In this paper, we developed a ridge regression estimator for Cox models and evaluated its performance through simulation studies. The method was then applied to a mortality study on a cohort from the Oak Ridge National Laboratory to evaluate the health risks associated with occupational radiation exposure.

2 Methods

2.1 Ridge estimator for linear and logistic regression models

The ridge estimator was originally defined for a linear regression model, $EY = X\beta$, as the estimator with minimum length among those that increase the sum of squared error by a small fixed amount:

$$\hat{\beta}^R = (X^T X + kI)^{-1} X^T X \hat{\beta}$$

where $\hat{\beta} = (X^T X)^{-1} X^T Y$ is the maximum likelihood estimator (MLE) for β (Hoerl and Kennard 1970a, b) and k is a constant, $k \geq 0$. The ridge estimator $\hat{\beta^R}$ minimizes the residual sum of squares subject to a constraint on the length of the estimated coefficient estimator (Gibbons 1981). Schafer et al. (1984) extended the ridge estimator to a logistic regression model, $logit(EY) = X\beta$, as follows:

$$\hat{\beta}^R = (X^T V X + kI)^{-1} X^T V X \hat{\beta}$$

where $\hat{\beta} = (X^T V X)^{-1} X^T V Y$ is the MLE for β and V is the covariance matrix of Y. Schaefer et al. have shown that $\hat{\beta^R}$ approximately minimizes the weighted sum of squared error (WSSE) with minimal length. Unlike the MLEs, the ridge estimators are not unbiased. In the presence of collinearity, the ridge estimator increases the bias a little but reduces the variance considerably, therefore it has a smaller MSE than the MLE.

2.2 Ridge estimator for Cox regression models

For a Cox proportional hazards model with a log-linear relative risk, $\lambda(t) = \lambda_0(t)e^{X\beta}$, Lustbader (1986) showed that the partial likelihood function is equivalent to the likelihood function of independently sampled Poisson random variables. Thus, the maximum likelihood estimator $\hat{\beta}$, obtained through maximizing the partial likelihood function, is also a solution to an iterative weighted least squares procedure of a linear regression (see the Appendix for details). The MLE can be derived as the following:

$$\hat{\beta} = (\hat{D}^T \hat{D})^{-1} \hat{D}^T \hat{U} \qquad (1)$$

where the matrices $\hat{D}$, $\hat{U}$ are functions of $\hat{\beta}$, the covariate matrix X, survival time t and its corresponding censoring indicator. We similarly define the ridge estimator $\hat{\beta}^R$ to be the estimator with minimum length among those that increase the WSSE by a small fixed amount:

$$\hat{\beta}^R = (\hat{D}^T \hat{D} + kI)^{-1} \hat{D}^T \hat{D} \hat{\beta} \qquad (2)$$

where k is a constant that needs to be chosen. The ridge estimator can thus be calculated based on the MLE and the estimated second derivative of the log partial likelihood function, which can be easily obtained using standard statistical software. There is no definitive rule for choosing k, but the general idea is to produce only a small increase in the WSSE. For logistic regression models, the most common choices of ridge parameters have been $1/\hat{\beta}^T \hat{\beta}$, $p/\hat{\beta}^T \hat{\beta}$, and $(p+1)/\hat{\beta}^T \hat{\beta}$ (Schafer et al. 1984; Smith et al. 1991) where p is the number of covariates in the regression. These three choices were investigated for Cox proportional hazards models in our simulation studies.

Unlike the MLEs for a linear or a generalized linear regression model, the maximum partial likelihood estimators are only asymptotically unbiased. Thus, the ridge estimators do not necessarily have more bias than the MLEs. Huang and Harrington (2002) showed that their penalized partial likelihood approach in fact provided some bias correction over the MLE for finite samples when there was collinearity in the covariate space. In this paper, we used simulation studies to evaluate the performance of the ridge estimators in terms of bias and MSE and compared those with the MLE.

The MLE maximizes the log-likelihood function and the ridge estimator minimizes the WSSE. Both the log-likelihood and the WSSE measure the degree of variation in the data explained by the model. It is also important to assess how well the model predicts future data. As there is no future data, the prediction process can be mimicked by crossvalidation: every observation is left out once and predicted by using all other observations. Verweij and Van Houwelingen (1993) proposed a crossvalidated partial likelihood (*cvl*) function to measure the predictive value of a Cox model. Here we used their measure to evaluate the predictive ability of the ridge estimator. Specifically, the cvl is defined as the follows:

$$cvl = \sum_{i=1}^{n} l_{Pi}(\hat{\beta^R_{-i}})$$

where $l_{Pi}(\beta) = l(\beta) - l_{P-i}(\beta)$ is the contribution of the ith individual to the log of the partial likelihood and $l_{P-i}(\beta)$ is the log partial likelihood when observation i is left out and $\hat{\beta^R_{-i}}$ is the ridge estimator for β when individual i is left out. The predictive abilities of the ridge estimator and the MLE were assessed and compared through simulation studies.

2.3 Confidence intervals in ridge regression

Crivelli et al. (1995) has showed that if the sample size is sufficiently large, the ridge estimator from a linear regression is asymptotically consistent when the ridge parameters are chosen to be $C/\hat{\beta}^T\hat{\beta}$ where C is a constant. We generalized their method to Cox models and estimated the variance of the ridge estimators by

$$\hat{Var}(\hat{\beta}^R) \approx (\hat{D}^T\hat{D} + kI)^{-1}\hat{D}^T\hat{D}(\hat{D}^T\hat{D} + kI)^{-1}.$$

Therefore, for a large sample, a $100(1-\alpha)\%$ approximate confidence interval for β_j is given by

$$\hat{\beta^R_j} + / - z_{1-\alpha/2}\sqrt{\hat{V^R}_{jj}} \tag{3}$$

where V^R_{jj} is the j, jth element of $Var(\hat{\beta}^R)$.

Verweij and Houwelingen (1994) suggested using the square root of the diagonal elements of $(\hat{D}^T\hat{D} + kI)^{-1}$ to measure the stability of the penalized estimates and referred to these as the pseudo-standard errors. In this paper, we constructed ridge regression confidence interval using both the above standard errors and compared their coverage probability (the probability of containing true β) and power (the probability of excluding 0) and also compare them with the MLE.

A nonparametric confidence interval was also constructed using the bootstrapping approach (Efron and Tibshirani 1993): the acceleration and bias-correction (BC_a) intervals which automatically corrects for bias. But a major disadvantage of the BC_a intervals is that a large number of bootstrap replication is required. Therefore, instead, we used approximate bootstrap intervals (ABC), a method of approximating the BC_a interval endpoints analytically without applying any Monte Carlo replications. The approximation is usually quite good (Efron and Tibshirani 1993). The program for computing the ABC intervals was downloaded from statlib at www.lib.stat.cmu.edu/S. The ABC intervals were compared with the intervals constructed using the methods described above in terms of coverage probability and power.

2.4 Comparison with the penalized likelihood approach

Verweij and Van Houwelingen (1994) developed a penalized likelihood estimator for the Cox regression model to increase the stability of the estimated regression coefficients. In their approach, a penalty function for the regression coefficient is subtracted from the partial log-likelihood as follows:

$$l_P^{\lambda}(\beta) = l_P(\beta) - \frac{1}{2}\lambda\beta^T A\beta$$

where A is a symmetric non-negative definite matrix. For a given λ, solution of this penalized partial likelihood function leads to the following estimator for β:

$$\hat{\beta}^{\lambda} = (\hat{D}^T\hat{D} + \lambda A)^{-1}\hat{D}^T\hat{D}\hat{\beta}.$$

Given λ, this is precisely the ridge estimate if A is chosen as the identity matrix. In Verweij and Van Houwelingen's approach, however, the weight parameter λ is determined by maximizing the predictive ability of the model measured by the *cvl* described above. Huang and Harrington (2002) proposed to estimate λ by minimizing the MSE for $x\beta$ and they have shown that their approach and Verweij and Van Houwelingen's approach are asymptotically equivalent. Apparently, their penalized estimates have the extra advantage of being the best predictors. However, not only is the penalized likelihood approach much more computationally intensive than the ridge estimation approach, it also ignores the source of variation due to the estimation of λ. As van Houwelingen (2001) pointed out, it is "not clear what the price is of estimating λ. It might be quite high". In this paper, we used simulation studies to compare the ridge estimator with Verweij and Van Houwelingen's penalized likelihood estimator with respect to finite sample properties.

3 Simulation study

A simulation study was conducted to evaluate the performance of the ridge estimator for Cox proportional hazards models. The simulation study was designed as follows: the size of the sample n was set to be 50. We assumed that the individuals were followed until death without any censoring. Therefore, there were 50 events in the sample. The number of covariates p was chosen to be four and six. A small sample size relative to the number of covariates was chosen since collinearity is more likely to affect parameter estimation for small samples. For large samples, collinearity is less of an issue. The parameters were set to be $(0.41, 0.24, -0.5, -0.2)$ for $p = 4$ and $(0.41, 0.24, -0.5, -0.2, -0.3, 0.4)$ for $p = 6$. The values of β's were chosen to be close to the parameter estimates in the application to the Oak Ridge data set. Continuous variables were generated from a multivariate normal distribution with mean 0 and variance 1. Pairwise correlations were set to be small ($\rho = 0.4$), moderate ($\rho = 0.6$) and heavy ($\rho = 0.8$). The baseline survival time (when all the covariates equal 0) was assumed to follow an exponential distribution with a 99% survival rate each year and a median survival time of 75 years.

In each simulated data set, we computed Huang and Harrington's measure of bias, the mean absolute proportion bias defined below:

$$\frac{1}{p}\sum_{i=1}^{p} \| \frac{\hat{\beta}_j^R}{\beta_j} - 1 \|$$

for the ridge estimator and substituted $\hat{\beta}_j^R$ with $\hat{\beta}_j$ for the MLE. We also computed the MSE averaged over p covariates and the *cvl* for both the MLE and the ridge estimator. Assuming β_1 is of primary interest, we computed the 95% ridge confidence interval for β_1 using the three different methods described in Sect. 2: (1) the standard error suggested by Crivelli et al. (1995); (2) the pseudo-standard error suggested by Verweij and Van Houwelingen (1994); (3) the nonparametric ABC interval obtained by bootstrapping. Because of the computational burden, we repeated the simulation 100 times and summarized the results in Tables 1 and 2 we present the average of the mean absolute proportion bias, the average of the MSE and the average of *cvl* for both the MLE and the ridge estimator. The performance of the ridge interval estimator for β is shown in Table 2. In Table 2, we present the coverage probability measured by the proportion of the confidence intervals including β_1 and the power measured by the proportion of the confidence intervals excluding 0 associated with the three confident intervals.

Table 1 indicates that the ridge estimator had smaller bias and MSE and higher *cvl* compared to the MLE: for example, when $\rho = 0.6$, the ridge estimator show approximately 15% reduction in bias, 30% reduction in MSE and 1.0–2.4 gain in cvl when $k = p/\hat{\beta}'\hat{\beta}$. The higher the ridge parameter k, the better the ridge estimator: for example, when $\rho = 0.6$ and $p = 4$, the reduction in bias increases from 7% to 17%, the reduction in MSE increases from 14% to 34% and the gain in *cvl* increases from 0.4 to 1.0 as k changes from $1/\hat{\beta}'\hat{\beta}$ to $(p + 1)/\hat{\beta}'\hat{\beta}$. Table 1 also indicates that the higher the collinearity (ρ), the better the ridge estimator compared to the MLE: for example, when ρ changed from 0.4 to 0.8 for $p = 6$, the reduction in bias increased from 17% to 27%, the reduction in MSE increased from 30% to 46% and the gain in *cvl* changed from 2.1 to 3.0 when $k = p/\hat{\beta}'\hat{\beta}$. The improvement in ridge estimators over the MLE is similar between $p = 4$ and $p = 6$, except that the gain in *cvl* is a bit higher for $p = 6$. In Fig. 1 the *cvl* from the ridge estimation was plotted against the *cvl*

Table 1 Comparison of bias, MSE and *cvl* between the MLE and the ridge estimators under various k for a Cox model with all continuous covariates (n=50)

β	ρ	Method	Bias[a] (rela. to MLE)	MSE[a] (rela. to MLE)	*cvl*[b] (diff from MLE)
(0.41,0.24,−0.5,−0.2)	0.4	MLE	0.62(1.00)	0.052(1.00)	−193.3 (0.0)
		$k = 1/\hat{\beta}'\hat{\beta}$	0.59(0.95)	0.046(0.88)	−192.9 (0.4)
		$k = p/\hat{\beta}'\hat{\beta}$	0.54(0.87)	0.037(0.72)	−192.5 (0.8)
		$k = (p+1)/\hat{\beta}'\hat{\beta}$	0.53(0.85)	0.037(0.72)	−192.4 (0.9)
	0.6	MLE	0.72(1.00)	0.073(1.00)	−195.1 (0.0)
		$k = 1/\hat{\beta}'\hat{\beta}$	0.67(0.93)	0.062(0.86)	−194.7 (0.4)
		$k = p/\hat{\beta}'\hat{\beta}$	0.61(0.85)	0.050(0.68)	−194.1 (1.0)
		$k = (p+1)/\hat{\beta}'\hat{\beta}$	0.60(0.83)	0.048(0.66)	−194.1 (1.0)
	0.8	MLE	1.02(1.00)	0.141(1.00)	−196.7 (0.0)
		$k = 1/\hat{\beta}'\hat{\beta}$	0.92(0.90)	0.111(0.80)	−196.2 (0.5)
		$k = p/\hat{\beta}'\hat{\beta}$	0.78(0.76)	0.077(0.54)	−195.5 (1.2)
		$k = (p+1)/\hat{\beta}'\hat{\beta}$	0.75(0.74)	0.072(0.52)	−195.4 (1.3)
[c](0.41,0.24,−0.5,−0.2,−0.3,0.2)	0.4	MLE	0.59(1.00)	0.054(1.00)	−193.6 (0.0)
		$k = 1/\hat{\beta}'\hat{\beta}$	0.56(0.95)	0.048(0.89)	−193.0 (0.6)
		$k = p/\hat{\beta}'\hat{\beta}$	0.49(0.83)	0.038(0.70)	−191.5 (2.1)
	0.6	MLE	0.71(1.00)	0.077(1.00)	−196.3 (0.0)
		$k = 1/\hat{\beta}'\hat{\beta}$	0.68(0.96)	0.068(0.88)	−195.6 (0.7)
		$k = p/\hat{\beta}'\hat{\beta}$	0.59(0.83)	0.053(0.69)	−193.9 (2.4)
	0.8	MLE	1.03(1.00)	0.156(1.00)	−199.0 (0.0)
		$k = 1/\hat{\beta}'\hat{\beta}$	0.94(0.91)	0.131(0.84)	−198.0 (1.0)
		$k = p/\hat{\beta}'\hat{\beta}$	0.75(0.73)	0.083(0.53)	−196.0 (3.0)

[a] The number in the parentheses is the ratio between the bias (MSE) of the ridge estimator and the bias of the MLE (MSE)
[b] The number in the parentheses is the difference between the *cvl* of the ridge estimator and the *cvl* of the MLE
[c] When $p = 6$, the performances of the ridge estimator under $k = p/\hat{\beta}'\hat{\beta}$ and $k = (p+1)/\hat{\beta}'\hat{\beta}$ are very close therefore only the ridge estimator with $k = p/\hat{\beta}'\hat{\beta}$ is reported

Table 2 Comparison of power and coverage probabilility between the MLE and the ridge estimator for β_1 under a Cox model with all continuous covariates

β	ρ	MLE		ridge	$se(\hat{\beta}_1^R)$		Pseudo-se$(\hat{\beta}_1^R)$		ABC Method	
		cover (%)	power (%)	k	cover (%)	power (%)	cover (%)	power (%)	cover (%)	power (%)
(0.41,0.24,−0.5,−0.2)	0.4	93	66	$1/\hat{\beta}'\hat{\beta}$	92	67	93	60	91	70
				$p/\hat{\beta}'\hat{\beta}$	84	63	87	55	89	72
				$(p+1)/\hat{\beta}'\hat{\beta}$	79	63	86	53	87	74
	0.6	95	47	$1/\hat{\beta}'\hat{\beta}$	92	49	94	44	94	41
				$p/\hat{\beta}'\hat{\beta}$	83	48	93	34	87	41
				$(p+1)/\hat{\beta}'\hat{\beta}$	75	48	91	29	85	40
	0.8	96	34	$1/\hat{\beta}'\hat{\beta}$	91	33	95	27	95	38
				$p/\hat{\beta}'\hat{\beta}$	75	31	93	22	87	36
				$(p+1)/\hat{\beta}'\hat{\beta}$	70	31	90	20	85	37
(0.41,0.24,−0.5,−0.2,−0.3,0.2)	0.4	92	55	$1/\hat{\beta}'\hat{\beta}$	92	56	94	54	92	59
				$p/\hat{\beta}'\hat{\beta}$	79	53	88	42	84	56
	0.6	95	47	$1/\hat{\beta}'\hat{\beta}$	92	49	94	44	94	41
				$p/\hat{\beta}'\hat{\beta}$	83	48	93	34	87	41
	0.8	92	26	$1/\hat{\beta}'\hat{\beta}$	93	26	97	25	97	28
				$p/\hat{\beta}'\hat{\beta}$	78	25	93	17	87	27

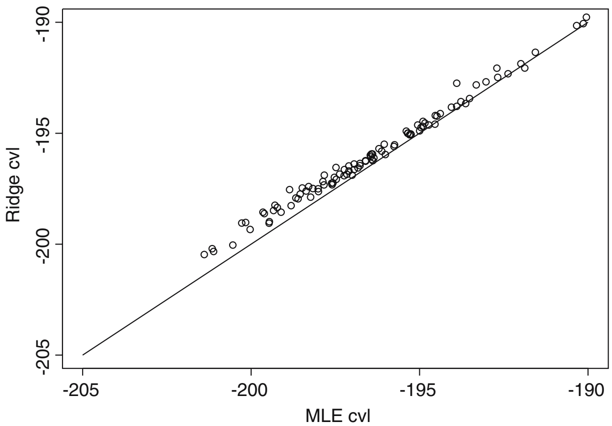

Fig. 1 Comparison of CVL between ridge estimator and MLE (line represents equal cvl)

from the MLE when $\rho = 0.8$, $p = 4$ and $k = 1/\hat{\beta}'\hat{\beta}$. Figure 1 shows that the *cvl* from the ridge regression is in general greater than that from the MLE. Table 2 shows that the ridge interval estimator based on the pseudo-standard error has the best coverage probability and the coverage probability is similar to that of the MLE either when k is small or when ρ is moderate or large. Neither the ridge interval estimator based on Crivelli's standard error nor the non-parametric ridge interval has good coverage probabilities but they have a slightly more power compared to the MLE.

We have also considered the case when the model has both continuous and categorical variables. After four continuous variables were generated from a multivariate normal distribution with the same parameters described above, we dichotomized two covariates into binary variables (0 or 1) with 30% and 50% chance of being 1. These proportions were chosen so that the distributions of the binary variables are close to what we observed in the application. The ridge analysis was applied and the results are summarized in Tables 3 and 4. As shown in Tables 3 and 4 our findings are quite consistent with what we observed in Table 1 and 2, except that the gain in power with ridge estimators is more apparent in Tables 3 and 4.

We increased the number of the simulations to 200 times in a few scenarios (not shown here) and the results remain the same; therefore, we conclude that these results are stable. Based on these findings, we conclude that ridge estimators are more precise and have greater predictive ability than the MLE in the presence of high collinearity. However, among the three ridge interval estimators, only the ridge interval estimator based on the pseudo-standard error has good coverage probability when ridge parameters are small or when collinearity is high. The other two ridge interval estimators do not have good coverage probabilities but have some gain in power compared to the MLE. Therefore, for interval estimation, we recommend the ridge interval estimate based on the pseudo-standard error; for statistical inference, we recommend the other two ridge interval estimates.

To compare the ridge estimator with the penalized likelihood estimator developed by Verweij and Van Houwelingen (1994), we computed the penalized likelihood

Table 3 Comparison of bias, MSE and *cvl* between the MLE and the ridge estimators under various k for a Cox model with two continuous and two binary covariates (n=50, $\beta_1 = log1.5$, $\beta_2 = 0.24$, $\beta_3 = -0.5$ and $\beta_4 = -0.2$)

ρ	method	bias (rela. to MLE)	MSE (rela.to MLE)	*cvl* (diff from MLE)
0.4	MLE	0.85(1.00)	0.111(1.00)	−193.4(0.0)
	$k = 1/\hat{\beta}'\hat{\beta}$	0.76(0.89)	0.091(0.82)	−192.9(0.5)
	$k = p/\hat{\beta}'\hat{\beta}$	0.65(0.76)	0.067(0.60)	−192.2(1.2)
	$k = (p+1)/\hat{\beta}'\hat{\beta}$	0.64(0.75)	0.066(0.59)	−192.1(1.3)
0.6	MLE	0.99(1.00)	0.140(1.00)	−193.5(0.0)
	$k = 1/\hat{\beta}'\hat{\beta}$	0.87(0.88)	0.112(0.80)	−193.0(0.5)
	$k = p/\hat{\beta}'\hat{\beta}$	0.73(0.74)	0.081(0.58)	−192.2(1.2)
	$k = (p+1)/\hat{\beta}'\hat{\beta}$	0.71(0.72)	0.077(0.55)	−192.2(1.2)
0.8	MLE	1.11(1.00)	0.178(1.00)	−194.2(0.0)
	$k = 1/\hat{\beta}'\hat{\beta}$	0.93(0.84)	0.132(0.74)	−193.5(0.7)
	$k = p/\hat{\beta}'\hat{\beta}$	0.74(0.67)	0.089(0.50)	−192.7(1.5)
	$k = (p+1)/\hat{\beta}'\hat{\beta}$	0.71(0.64)	0.084(0.47)	−192.6(1.6)

Table 4 Comparison of power and coverage probabilility between the MLE and ridge estimator for β_1 under a Cox model with two continuous and two binary covariates (n=50, $\beta_1 = log1.5$, $\beta_2 = 0.24$, $\beta_3 = -0.5$ and $\beta_4 = -0.2$)

ρ	MLE		ridge	se($\hat{\beta}_1^R$)		pseudo-se($\hat{\beta}_1^R$)		ABC Method	
	cover (%)	power (%)	k	cover (%)	power (%)	cover (%)	power (%)	cover (%)	power (%)
0.4	95	77	$1/\hat{\beta}'\hat{\beta}$	93	76	94	73	89	76
			$p/\hat{\beta}'\hat{\beta}$	87	75	92	64	88	81
			$(p+1)/\hat{\beta}'\hat{\beta}$	87	75	91	61	89	80
0.6	96	56	$1/\hat{\beta}'\hat{\beta}$	94	59	95	55	94	55
			$p/\hat{\beta}'\hat{\beta}$	89	64	94	45	92	62
			$(p+1)/\hat{\beta}'\hat{\beta}$	88	64	93	44	91	65
0.8	95	40	$1/\hat{\beta}'\hat{\beta}$	91	41	96	37	94	45
			$p/\hat{\beta}'\hat{\beta}$	87	46	93	26	93	61
			$(p+1)/\hat{\beta}'\hat{\beta}$	85	51	93	23	89	63

estimator for $p = 4$ and $\rho = 0.8$. The results are summarized in Table 5. In the penalized likelihood approach, the weight parameter λ is chosen through maximizing the *cvl*. Table 5 shows that the estimated λ is higher than the chosen ridge parameters. The penalized likelihood estimate has the highest averaged *cvl*, similar averaged bias and MSE compared to the ridge estimates when $k = p/\hat{\beta}'\hat{\beta}$ and $k = (p+1)/\hat{\beta}'\hat{\beta}$. However, the penalized likelihood interval estimate has a much lower coverage probability and a lower power. Therefore, if the primary interest is to obtain a point estimate with highest predictive ability, the penalized likelihood estimator may be preferred.

Table 5 Comparison between the penalized likelihood estimator and the ridge estimator in a Cox model with four continuous covariates(n=50, $\beta_1 = log1.5$, $\beta_2 = 0.24$, $\beta_3 = -0.5$ and $\beta_4 = -0.2$)

	$\bar{\hat{\lambda}}$	*cvl*	bias	MSE	cover[1](%) for β_1	power[a] (%) for β_1
Penal.	16	−194.5	0.77	0.072	60	19
$k = 1/\hat{\beta}'\hat{\beta}$	3	−196.2	0.92	0.111	95	27
$k = p/\hat{\beta}'\hat{\beta}$	7	−195.5	0.78	0.077	93	22
$k = (p+1)/\hat{\beta}'\hat{\beta}$	10	−195.4	0.75	0.072	90	20

[a]The coverage and power associated with the ridge estimator is based on the pseudo-standard error. Results for the ridge estimator were obtained from Table 1.

But if the primary interest is to obtain an accurate and precise point as well as interval estimate, the ridge estimator will be preferred. Since a ridge estimate is much less computationally intensive to obtain, even when predictive ability is of primary interest, we still recommend that one estimates the ridge estimator to compare to the penalized likelihood estimator.

4 An application

We applied the proposed ridge estimator for the Cox model to a data set from the Oak Ridge National Laboratory (ORNL). ORNL is a US Department of Energy research and development facility that began operation in Oak Ridge, Tenn, in 1943. It is one of several facilities included in a large follow-up study of the health and mortality of workers at DOE facilities (Gilbert et al. 1993b; Shore 1990). The original study population consists of 8,318 white males hired at ORNL between the opening of the facility in January, 1943 and the end of 1972. The men all worked for at least thirty days, and there is no record indicating they had been employed at any other facility during the time period of interest. The main objective is to evaluate the health effects of occupational radiation exposure (Wing et al. 1991). In this paper, we considered a subset of 2,134 workers who entered the factory after 1957 and had worked there for at least a year. The reason that only the workers who started after 1957 were considered is to ensure the quality of dose estimates: in earlier years the dose estimates were subject to heavy censoring due to the minimum detection limit of the dosimeter (Xue et al. 2004). The radiation exposure varies largely between subjects as well as within subjects over time. Figure 2 shows a history of annual radiation exposures for a random sample of 100 workers. Because of the non-constant exposure rate, cumulative exposure has been used to measure radiation exposure (Wing et al. 1991, 1993; Frome et al. 1997; Cardis et al. 2005). It takes many years for the radiation exposure to manifest its carcinogenic effect. Previous analyses have shown that a 20-year lag for cumulative radiation dose fits the data better than shorter lag assumptions (Wing et al. 1993; Frome et al. 1997; Cardis et al. 2005). Therefore, a 20-year lag was used in this paper. A Cox model was used to evaluate the association of radiation exposure with all-cause mortality, in which the cumulative radiation exposure was treated as a time-dependent variable. The following variables were considered

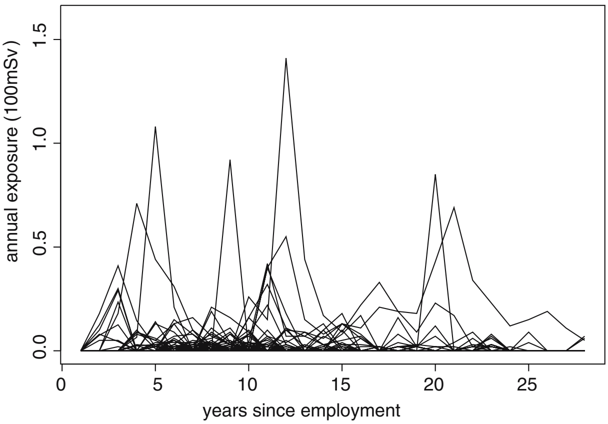

Fig. 2 Exposure History for a Random Selected 100 Subjects from the ORNL Subset

as potential confounders in the dose response analysis: (1) age at exposure divided into three categories: below 30, 30–39 and above 40; (2) birth year divided into three categories: before 1930, 1930–1939 and 1940 or after; These two variables have been conventionally treated as categorical variables instead of continuous ones in the radiation epidemiological literature as from a biological perspective it is too restrictive to assume a linear association between these two variables and risk for mortality (Gilbert 1989; Gilbert et al. 1993a,b; Frome et al. 1997) (3) socio-economic status for which paycode was used as a surrogate marker: monthly paid vs. non-monthly paid with the monthly paid to be a higher socioeconomic category; (4)number of years exposed to radiation exposure as a continuous time-dependent variable; (5) active worker status (currently employed) was also included as a time-dependent binary variable to take into account the healthy worker effect because workers who continued employment, and consequently remained exposed, tend to be healthier. Data on smoking, chemical exposures, medical exposures to ionizing radiation, and cancer mortality were not available. A Cox model was applied to the data set and the maximum likelihood estimates were summarized in Table 6. Table 6 shows that every additional $10mSV$ dose of radiation exposure increases the risk of death by 4%, however, the effect is not statistically significant. The distribution of radiation exposure was skewed with most of them being zero or very small doses and a few with large doses. The cumulative dose difference between the 90th percentile and the 10th percentile is about $50mSv$ (Wing et al. 1991), which results in an estimated relative risk for mortality of 1.2. Table 6 also shows that age below 30 at initial exposure is associated with the highest risk for death. The risk decreases dramatically if exposure occurred at a later age. Birth before 1930 is associated with the highest risk and birth after 1930 has a much lower risk. High social economic status (monthly paid) is a protective factor. Active worker (currently employed) is associated with less risk of death than nonactive workers with borderline significance. Risk increases 2% for every additional year of exposure given the same level of cumulative exposure. These findings are consistent with what has been seen in the literature (Wing et al. 1993; Frome et al. 1997; Cardis et al. 2005).

Table 6 Estimation of the Cox Model for all-cause mortality for the ORNL subset

Variable		est	se	p-value	haz. ratio	CI
Cumdose(100mSV)		0.407	0.514	0.429	1.502	(0.548,4.114)
Birth year	<1930	ref				
	[1930, 1939]	−1.963	0.440	<.001	0.140	(0.059,0.333)
	≥1940	−4.374	0.668	<.001	0.013	(0.003,0.047)
Age at exp	<30	ref				
	[30, 39]	−1.040	0.414	0.012	0.353	(0.157,0.796)
	≥40	−1.197	0.531	0.024	0.302	(0.107,0.855)
Paycode	weekly	ref				
	monthly	−0.642	0.248	0.010	0.526	(0.323,0.857)
Current employ	no	ref				
	yes	−0.397	0.236	0.092	0.672	(0.423,1.067)
Years exposed		0.022	0.003	<.001	1.022	(1.017,1.027)

However, the potential collinearity issue in occupational radiation exposure studies has not been addressed previously. Years of exposure is likely to be positively correlated with cumulative exposure. Currently-employed status also tends to be correlated with cumulative dose and years of exposure. The sample correlation between variables was calculated for each year between 1977 and 1984 (since a lag of twenty years was used for the exposure variable) and averaged over these years. Table 7 indicates that there is a strong correlation between current employment status and years of exposure (0.81), although the primary exposure variable, the cumulative radiation exposure, is not highly associated with any other covariates. There are a total of 103 deaths in this sample. With eight covariates in the model, collinearity can have a big impact on the estimators. We therefore applied the ridge regression and summarized the results in Table 8. Table 8 shows that the ridge estimator for the relative risk associated with cumulative radiation exposure is very close to the MLE, especially when the ridge parameter is $1/\hat{\beta}'\hat{\beta}$. As the ridge parameter increases, the relative risk estimate is reduced by up to 6% only. The high consistency in the estimates of the relative risk associated with radiation exposure between the ridge estimation and the MLE may result from the fact that the cumulative radiation exposure is not highly associated with any of the other covariates. The ridge confidence intervals are narrower but all include 1. The *cvl* is very close to that of the MLE and is even slightly higher for larger ridge parameters, indicating that in this case the ridge estimation does not offer an improvement in predictive ability over the MLE. In summary, we conclude that the previous finding of no association between occupational radiation exposure and all-cause mortality remains valid after adjusting for collinearity between covariates (Table 9).

5 Conclusion and discussion

In this paper, we proposed to extend the ridge regression method derived for linear regression and logistic regression for the treatment of collinearity to the analysis

Table 7 Sample correlation between covariates in the Cox Model on all-cause mortality for an ORNL subcohort

Variables	Cumdose	Birth year [1930, 1939]	Birth year ≥ 1940	Age exp [30, 39]	Age exp ≥ 40	Monthly paycode	Current employer	Years exposed
Cumdose	1.00	0.04	−0.13	−0.01	0.03	−0.03	0.02	0.10
Birth year [1930, 1939]		1.00	−0.64	0.00	−0.24	0.18	0.08	0.12
Birth year ≥1940			1.00	−0.37	−0.19	−0.24	−0.09	−0.22
Age exp [30, 39]				1.00	−0.16	0.13	0.10	0.14
Age exp ≥40					1.00	0.02	−0.06	0.03
Monthly paycode						1.00	−0.05	−0.03
Current employer							1.00	0.81
Years exposed								1.00

Table 8 Ridge estimation for hazard ratio associated with cumulative radiation exposure for the ORNL subset

ridge para k	ridge est. haz. ratio	CI based on se($\hat{\beta}_1^R$)	CI based on pseudo-se($\hat{\beta}_1^R$)	ABC CI
$1/\hat{\beta}'\hat{\beta}$	1.488	(0.549,4.035)	(0.546, 4.056)	(0.587, 2.889)
$p/\hat{\beta}'\hat{\beta}$	1.412	(0.556,3.586)	(0.536, 3.722)	(0.666, 2.448)
$(p+1)/\hat{\beta}'\hat{\beta}$	1.404	(0.558,3.533)	(0.535, 3.682)	(0.670, 2.396)

Table 9 *cvl* for MLE and Ridge Estimation for the ORNL subset

	Ridge Estimator		
MLE	$k = 1/\hat{\beta}'\hat{\beta}$	$k = p/\hat{\beta}'\hat{\beta}$	$k = (p+1)/\hat{\beta}'\hat{\beta}$
−320.6	−320.6	−321.1	−321.2

of time to event data. The ridge regression estimator was based on a Cox proportional hazards model. Ridge regression estimates can be easily obtained using standard statistical software. Simulation studies showed that the ridge estimator is more precise and accurate and has better predictive ability compared to the MLE, especially when the collinearity is high in a finite sample. The advantages are greater if a larger ridge parameter is used or if the collinearity is high. Several methods were proposed to estimate the stability of the ridge estimates. Simulation studies showed that the ridge interval estimator based on pseudo-standard errors gives good coverage probability for smaller ridge parameters. Intervals based on either Crivelli's method or the nonparametric bootstrap method do not have good coverage probabilities but have a small to moderate gain in power compared to the MLE. Choices of the ridge parameter and ridge interval estimators depend on one's primary interest: if the primary interest is obtain a precise and accurate point estimate, a larger ridge parameter should be used; if the primary interest is to obtain an interval estimate with good coverage probability, a small ridge parameter and the interval based on pseudo-standard error should be used; if the primary interest is to make statistical inference on exposure-disease association, the other two ridge intervals should be used.

Ridge regression is a special case of the penalized likelihood approach proposed by Verweij and van Houwelingen (1994) and Huang and Harrington (2002). Ridge regression is computationally simpler than a regular penalized likelihood approach and also offers the flexibility of being able to choose the ridge parameter based on one's primary interest. Although the penalized likelihood estimator can have a higher *cvl* than a ridge estimator, our simulations showed that it can also have smaller coverage and lower power. Therefore, unless predictability of the model is the primary interest, we recommend the ridge regression.

We applied the method to a mortality study from ORNL, in which covariates such as radiation exposure, years exposure and active worker status tend to be highly correlated with each other. Ridge estimates for the hazard ratio associated with the primary exposure variable, the cumulative radiation exposure are slightly smaller than the MLE with narrower confidence intervals. The similarities between the ridge estimates and the MLE may result from lack of high correlations between the cumulative radiation exposure and other covariates. The conclusion that occupational radiation exposure is not related to risk of death remains the same after taking into account collinearity.

The validity of ridge regression depends on the validity of the maximum likelihood estimation. Our simulations showed that with 50 events and 6 covariates in the model, the maximum likelihood estimation does not have difficult to converge. However, when the number of covariates is larger than number of events, the MLE is ill-defined so is the ridge regression. Thus, the ridge regression method is not a solution to the analysis of high dimensional data. Methods to reduce the dimension of the data such as factor analysis have to be used instead.

An alternative to ridge regression to control for collinearity is a method called the Lasso, for "least absolute shrinkage and selection operator" proposed by Tibshirani (1996, 1997). The Lasso estimator maximizes the likelihood function (or the partial likelihood function) subject to the sum of the absolute value of the coefficients being than less than a constant. Because of the nature of this constraint it tends to produce some coefficients that are exactly 0 and hence gives simpler models than approaches such as ridge regressions which retain all the covariates. We plan to compare the Lasso estimator with the ridge estimator in future studies.

Acknowledgements The authors are grateful for the valuable comments made by the associate editor and the two referees. This work was supported by NIOSH/CDC grant 1R01 OH008087.

Appendix

Let n denote the number of uncensored survival times in the study sample. Assume there are no ties and let i_0 represent the uncensored survival time in the ith risk set. The partial likelihood L_P can be written as follows:

$$L_P = \prod_{i=1}^{n} \frac{\rho(\beta, x_{i0})}{\sum_{l=0}^{m_i} \rho(\beta, x_{il})}$$

where the relative risk $\rho(\beta, x) = exp(x\beta)$ for a log-linear relative risk model and $m_i + 1$ is the size of the ith risk set. Lustbader (1986) has shown that this likelihood is equivalent to the likelihood function of realizations of independent Poisson random variables Y_{ij} where $Y_{i0} = 1$ and $Y_{ij} = 0$ for $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$ and the expected value of Y_{ij} is

$$E(Y_{ij}) = P_{ij} = \frac{\rho(\beta, x_{ij})}{\sum_{l=0}^{m_i} \rho(\beta, x_{il})}$$

for $j = 0, 1, \ldots, m_i$. The Poisson view of the partial likelihood permits the straightforward extension of least square estimation to relative risk regression models. Lustbader (1986) has shown that a Poisson regression model can be reformulated as a linear regression model with a pseudo outcome variable

$$U = D\beta + e. \tag{4}$$

The row of the design matrix D (D has dimension $\sum(m_i + 1) * p$) is given by

$$\sqrt{P_{ij}}(\kappa_{ij} - \overline{\kappa_i})$$

where $\kappa_{ij} = \partial log\rho(\beta, x_{ij})/\partial(\beta)$ and $\overline{\kappa_i} = \sum_{l=0}^{m_i} P_{il}\kappa_{il}$. Under a log-linear relative risk model, $\kappa_{ij} = x_{ij}$. Note that the design matrix D is equivalent to the covariate matrix in a linear regression, except that the input data is the distance of κ_{ij} from its weighted risk set mean $\overline{\kappa_i}$. The elements of the residual term e are:

$$e_{ij} = \frac{y_{ij} - P_{ij}}{\sqrt{P_{ij}}}$$

for $j = 0, \ldots, m_i$ and $i = 1, \ldots, n$. The linear regression reformulation in (4) naturally gives rise to (1) and the ridge estimator defined in (2), where the MLE $\hat{\beta}$ is a solution to an iterative weighted least squares procedure of (4). Since

$$-\frac{\partial^2 log(L_p)}{\partial \beta^2}|_{\beta=\hat{\beta}} = \hat{D}^T \hat{D},$$

$\hat{D}^T \hat{D}$ is the estimated information matrix for the Cox model.

References

1. Barker L, Brown C (2001) Logistic regression when binary predictor variables are highly correlated. Stat Med 20:1431–1442
2. Cardis E, Gilbert E, Carpenter L et al (1995) Effects of low doses and low dose rates of external ionizing radiation: cancer mortality among nuclear industry workers in three countries. Radiat Res 142:117–132
3. Cardis E, Vrijheid M, Blettner M, Gilbert E, Hakama M, Hill C, Howe G, Kaldor J, Muirhead CR, Schubauer-Berigan M., Yoshimura T., Bermann F, Cowper G, Fix J, Hacker C et al (2005) Risk of cancer after low doses of ionizing radiation–retrospective cohort study in 15 countries. Br Med J 331:77–81
4. Crivelli A, Firinguetti L, Montano R, Munoz M (1995) Confidence intervals in ridge regression by bootstrapping the dependent variable: a simulation study. Commun Statist Simula 24:631–652
5. Cox DR (1972) Regression models and life tables (with discussion). J Royal Stat Soc Series B 34:187–220
6. Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman & Hall
7. Frome EL, Cragle DL, Watkins JP, Wing S, Shy CM, Tankersley WG, West CM (1997) A mortality study of employees of the nuclear industry in oak ridge, tennessee. Radiat Res 148:64–80
8. Gibbons DG (1981) A simulation study of some ridge estimators. J Am Stat Assoc 76:131–139
9. Gilbert ES (1989) Issues in analyzing the effects of occupational exposure to low levels of radiation. Stat Med 8:173–187
10. Gilbert ES, Omohundro E, Buchanan J, Holter N (1993a) Mortality of workers at the hanford site: 1945-1986. Health Phy 64:577–590
11. Gilbert ES, Cragle D, Wiggs L (1993b) Updated analyses of combined mortality data for workers at the hanford site, oak ridge national laboratory, and rocky flats weapons plant. Radiat Res 136:408–421
12. Huang J, Harrington D (2002) Penalized partial likelihood regression for right censored data with bootstrap selection of the penalty parameter. Biometrics 58:781–791
13. Hoerl AE, Kennard RW (1970a) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12:55–67
14. Hoerl AE, Kennard RW (1970b) Ridge regression: some simulations. Commun Stat 4(2):1105–1123
15. Holford TR, Zheng T, Mayne S, Zahm SH, Tessari JD, Boyle P (2000) Joint effects of nine polychlorinated biphenyl (PCB) congeners on breast cancer risk. Int J Epidemiol 29:975–982
16. Lagakos SW (1988) Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. Stat Med 7:257–274
17. Lustbader ED (1986) Relative risk regression diagnosis. In: Moolgavkar SH, Prentice RL (eds) Modern statistical methods in chronic disease epidemiology. SIAM, Philadelphia
18. Marsh JL, Hutton JL, Binks K (2002) Removal of radiation dose response effects: an example of over-matching. Brit. Med. J 325:327–330
19. Schafer RL, Roi LD, Wolfe RA (1984) A ridge logistic estimator. Commun Stat Theory Methods 13(1):99–113
20. Smith KR, Slattery ML, French TK (1991) Collinear nutrients and the risk of colon cancer. J Clin Epidemiol 44:715–723
21. Shore RE (1990) Occupational radiation studies: status, problems and prospects. Health Phys 59:63–68
22. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Royal Stat Soc Series B 58:267–288
23. Tibshirani R (1997) The lasso method for variable selection in the cox model. Stat Med 16:385–395
24. Van Houwelingen JC (2001) Shrinkage and penalized likelihood as methods to improve predictive accuracy. Stat Neerlandica 55:17–34

25. Verweij PJM , van Houwelingen HC (1994) Penalized likelihood in cox regression. Stat Med 13:2427–2436
26. Verweij PJM , van Houwelingen HC (1993) Cross-validation in survival analysis. Stat Med 12:2305–2314
27. Wing S, Shy CM, Wood JL, Wolf S, Cragle DL, Frome EL (1991) Mortality among workers at oak ridge national laboratory, evidence of radiation effects in follow-up through 1984. J Am Med Assoc 265:1397–1402
28. Wing S, Shy CM, Wood JL, Wolf S, Cragle DL, Tankersley W, Frome EL (1993) Job factors, radiation and cancer mortality at oak ridge national laboratory: follow up through 1984. Am J Indust Med 23:265–279
29. Wing S, Richardson DB (1999) Greater sensitivity to ionizing radiation at older age: follow-up of workers at oak ridge national laboratory through 1990. Int J Epidemiol 28: 428–436
30. Xue X, Shore RE, Ye X, Kim MY (2004) Estimating the dose response relationship for occupational radiation exposure measured with minimum detection level. Health Phys 87:0397–404