

5-1-2005

Bias Of The Cox Model Hazard Ratio

Inger Persson

TFS Trial Form Support AB, Stockholm, Sweden

Harry Khamis

Wright State University

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Persson, Inger and Khamis, Harry (2005) "Bias Of The Cox Model Hazard Ratio," *Journal of Modern Applied Statistical Methods*: Vol. 4 : Iss. 1 , Article 10.

DOI: 10.22237/jmasm/1114906200

Available at: <http://digitalcommons.wayne.edu/jmasm/vol4/iss1/10>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Bias Of The Cox Model Hazard Ratio

Inger Persson
TFS Trial Form Support AB
Stockholm, Sweden

Harry Khamis
Statistical Consulting Center
Wright State University

The hazard ratio estimated with the Cox model is investigated under proportional and five forms of nonproportional hazards. Results indicate that the highest bias occurs for diverging hazards with early censoring, and for increasing and crossing hazards under a high censoring rate.

Key words: censoring proportion, proportional hazards, random censoring, survival analysis, type I censoring

Introduction

In recent decades, survival analysis techniques have been extended far beyond the medical, biomedical, and reliability research areas to fields such as engineering, criminology, sociology, marketing, insurance, economics, etc. The study of survival data has previously focused on predicting the probability of response, survival, or mean lifetime, and comparing the survival distributions. More recently, the identification of risk and/or prognostic factors related to response, survival, and the development of a certain condition has become equally important (Lee, 1992).

Conventional statistical methods are not adequate to analyze survival data because some observations are censored, i.e., for some observations there is incomplete information about the time to the event of interest. A common type of censoring in practice is Type I censoring, where the event of interest is observed only if it occurs prior to some pre-

specified time, such as the closing of the study or the end of the follow-up. The most common approach for modeling covariate effects in survival data uses the *Cox Proportional Hazards Regression Model* (Cox, 1972), which takes into account the effect of censored observations. As the name indicates, the Cox model relies on the assumption of proportional hazards, i.e., the assumption that the effect of a given covariate does not change over time. If this assumption is violated, then the Cox model is invalid and results deriving from the model may be erroneous.

A great number of procedures, both numerical and graphical, for assessing the validity of the proportional hazards assumption have been proposed over the years. Some of the procedures require partitioning of failure time, some require categorization of covariates, some include a spline function, and some can be applied to the untransformed data set.

However, no method is known to be definitively better than the others in determining nonproportionality. Some authors recommended using numerical tests, e.g., Hosmer and Lemeshow (1999). Others recommended graphical procedures, because they believe that the proportional hazards assumption only approximates the correct model for a covariate and that any formal test, based on a large enough sample, will reject the null hypothesis of proportionality (Klein & Moeschberger, 1997, p. 354).

Power studies to compare some numerical tests have been performed; see, e.g.,

Harry Khamis, Statistical Consulting Center
Wright State University, Dayton, OH.
45435. Email him: harry.khamis@wright.edu.

Ng'andu, 1997; Quantin, et al., 1996; Song & Lee, 2000, and Persson, 2002. The goal of this article is to assess the bias of the Cox model estimate of the hazard ratio under different censoring rates, sample sizes, types of nonproportionality, and types of censoring. The second section reviews the Cox regression model and the proportional hazards assumption. The average hazard ratio, the principal criterion against which the Cox model estimates are compared, is described in the third section. The fourth section presents the simulation strategy. The results and conclusions are given in the remaining two sections.

Cox proportional hazards model

A central quantity in the Cox regression model is the hazard function, or the hazard rate, defined by:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t},$$

where T is the random variable under study: time until the event of interest occurs. Thus, for small Δt , $\lambda(t)\Delta t$ is approximately the conditional probability that the event of interest occurs in the interval $[t, t + \Delta t]$, given that it has not occurred before time t .

There are many general shapes for the hazard rate; the only restriction is $\lambda(t) \geq 0$. Models with increasing hazard rates may arise when there is natural aging or wear. Decreasing hazard functions are less common, but may occur when there is a very early likelihood of failure, such as in certain types of electronic devices or in patients experiencing certain types of transplants.

A bathtub-shaped hazard is appropriate in populations followed from birth. During an early period deaths result, primarily from infant diseases, after which the death rate stabilizes, followed by an increasing hazard rate due to the natural aging process. Finally, if the hazard rate is increasing early and eventually begins declining, then the hazard is termed "hump-shaped." This type of hazard rate is often used in modeling survival after successful surgery, where there is an initial increase in risk due to infection or other complications just after the

procedure followed by a steady decline in risk as the patient recovers (see, e.g., Kline & Moeschberger, 1997).

In the Cox model, the relation between the distribution of event time and the covariates \mathbf{z} (a $p \times 1$ vector) is described in terms of the hazard rate for an individual at time t :

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}), \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard rate, an unknown (arbitrary) function giving the value of the hazard function for the standard set of conditions $\mathbf{z} = \mathbf{0}$, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters. The partial likelihood estimate of $\boldsymbol{\beta}$ is asymptotically consistent (Andersen & Gill, 1982; Cox, 1975, and Tsiatis, 1981).

The ratio of the hazard functions for two individuals with covariate values \mathbf{z} and \mathbf{z}^* is $\lambda(t, \mathbf{z})/\lambda(t, \mathbf{z}^*) = \exp[\boldsymbol{\beta}'(\mathbf{z} - \mathbf{z}^*)]$, an expression that does not depend on t . Thus, the hazard functions are proportional over time. The factor $\exp(\boldsymbol{\beta}'\mathbf{z})$ describes the hazard ratio for an individual with covariates \mathbf{z} relative to the hazard at a standard $\mathbf{z} = \mathbf{0}$. The usual interpretation of the hazard ratio, $\exp(\boldsymbol{\beta}'\mathbf{z})$, requires that (1) holds. There is no clear interpretation if the hazards are not proportional.

Of principal interest in a Cox regression analysis is to determine whether a given covariate influences survival, i.e. to estimate the hazard ratio for that covariate. The behavior of the hazard ratio estimated with the Cox model when the underlying assumption of proportional hazards is false (i.e., when the hazards are not proportional) is investigated in this paper. To assess the Cox estimates under nonproportional hazards, the estimates are compared to an exact calculation of the geometric average of the hazard ratio described in the next section. An average hazard ratio does not reflect the truth exactly since the hazard ratio is changing with time when the proportionality assumption is not in force. However, it can provide an approximate standard against which to compare the Cox model estimates. Because the estimation of the hazard ratio from the Cox model cannot be done analytically (Klein & Moeschberger, 1997), the comparison is made by simulations.

Average hazard ratio

The average hazard ratio (AHR) is defined as (Kalbfleisch & Prentice, 1981):

$$\theta(W) = - \int_0^{\infty} [\lambda_1(t)/\lambda_2(t)]dW(t), \quad (2)$$

where $\lambda_1(t)$ and $\lambda_2(t)$ are the hazard functions of two groups and $W(t)$ is a survivor or weighting function. The weight function can be chosen to reflect the relative importance attached to hazard ratios in different time periods. Here, $W(t)$ depends on the general shape of the failure time distribution and is defined as $W(t) = S_1^\varepsilon(t)S_2^\varepsilon(t)$, where $S_1(t)$ and $S_2(t)$ are the survivor functions (i.e., one minus the cumulative distribution function) for the two groups, and $\varepsilon > 0$. The value $\varepsilon = 1/2$ weights the hazard ratio at time t according to the geometric average of the two survivor functions. Values of $\varepsilon > 1/2$ will assign greater weight to the early times while $\varepsilon < 1/2$ assigns greater weight to later times. Here, $\varepsilon = 1/2$ will be used.

For Weibull distributed lifetimes with scale parameter α and shape parameter γ , the survival function is $S(t) = \exp[-(\alpha t)^\gamma]$ and the AHR estimator (2) can be written

$$\theta(W) = - \int_0^{\infty} [(\gamma_1 \alpha_1^{\gamma_1}) / (\gamma_2 \alpha_2^{\gamma_2})] d\{\exp[-1/2((\alpha_1 t)^{\gamma_1} + (\alpha_2 t)^{\gamma_2})]\}.$$

When the parametric forms of the survivor functions are unknown, the AHR (2) can still be used; in this case, the Kaplan-Meier product-limit estimates for the two groups are used as the survivor functions (Kaplan & Meier, 1958). However, (2) then only holds for uncensored data. The AHR function for censored data can be found in Kalbfleisch and Prentice, 1981.

Methodology

Simulation strategy

The hazard ratio estimates from the Cox model are evaluated under six scenarios:

- (1) proportional hazards, (2) increasing hazards,
- (3) decreasing hazards, (4) crossing hazards,

- (5) diverging hazards, and (6) converging hazards. The AHR is compared in the two-sample case, corresponding to two groups with different hazard functions.

Equal sample sizes of 30, 50, and 100 observations per group are used along with average censoring proportions of 10, 25, and 50 percent. Type I censoring is used along with early and late censoring. The number of repetitions used in each simulation is 10,000. For a given sample size, censoring proportion, and type of censoring (random, early, late), the mean Cox estimate is calculated for all scenarios except converging hazards. Because of the asymmetry in the distribution of values in the case of converging hazards, the median estimate is used. For interpretation purposes, the percent bias of the mean or median Cox estimate relative to the AHR is reported in tables.

For the case of random censoring, random samples of survival times t_s are generated from the Weibull distribution. The hazard function for the Weibull distribution is $\lambda(t) = \alpha\gamma(\alpha t)^{\gamma-1}$. The censoring times t_c are generated from the exponential distribution with hazard function $\lambda(t) = \beta$, where the value of β is adjusted to achieve the desired censoring proportions. The time on study t is defined as:

$$t = \begin{cases} t_s & \text{if } t_s \leq t_c \\ t_c & \text{if } t_s > t_c \end{cases}$$

The event indicator is denoted by d :

$$d = \begin{cases} 0, & \text{if the observation is censored} \\ 1, & \text{if the event has occurred} \end{cases}$$

For early censoring, a percentage of the lifetimes are randomly chosen and multiplied by a random number generated from the uniform distribution. The percentage chosen is the same as the censoring proportion. The parameters of the uniform distribution are chosen so that the censoring times are short in order to achieve the effect of early censoring. For late censoring, a percentage of the longest lifetimes are chosen; this percentage is slightly larger than the censoring proportion. Of those lifetimes, a percentage corresponding to the censoring time

is the lifetime, t_s , minus a random number generated from the uniform distribution. The parameters of the uniform distribution are now chosen so that the random numbers are relatively small in order to achieve the effect of late censoring.

Results

For each of the six scenarios concerning the hazard rates of the two groups, comparisons of the estimated hazards ratio from the Cox model to the AHR is made for random, early, and late censoring and for selected sample sizes and censoring rates. The comparison is made based on the percent difference (bias) between the average Cox hazard ratio estimate and the AHR; $[(\text{average Cox estimate} - \text{AHR})/\text{AHR}] \times 100$.

Proportional Hazards

Survival times are generated from the Weibull distribution where $\gamma=1$, $\alpha=1$ for group 1, and $\gamma=1$, $\alpha=2$ for group 2. The AHR is 2.0 for this situation. The percent of the bias for the mean Cox model estimate relative to the AHR is given in Table 1.

Under proportional hazards, the Cox model is correct. So, the estimated hazard ratio from the Cox model should be close to 2.0 in all cases. Table 1 reveals that the Cox estimate is slightly biased. This bias grows with decreasing sample size or increasing censoring proportion. Early censoring produces a more biased estimate than random or late censoring, especially for high censoring proportions.

Increasing Hazards

Survival times are generated from the Weibull distribution where $\gamma=1.5$, $\alpha=2$ for group 1, and $\gamma=2$, $\alpha=2$ for group 2. The AHR is 1.2 for this situation. The percent of the bias for the mean Cox model estimate relative to the AHR is given in Table 2.

The Cox estimates fall below the AHR for increasing hazards. The estimates closest to the AHR correspond to early censoring; these estimates are relatively stable regardless of censoring proportion or sample size. For random and late censoring the estimate decreases (higher bias) with increasing censoring proportion but remains stable relative to sample size. For early

censoring the estimate is generally unbiased regardless of sample size or censoring proportion.

Decreasing Hazards

Survival times are generated from the Weibull distribution where $\gamma=0.9$, $\alpha=1$ for group 1, and $\gamma=0.75$, $\alpha=3$ for group 2. The AHR is 0.44 for this situation. The percent of the bias for the mean Cox model estimate relative to the AHR is given in Table 3.

The Cox estimates fall below the AHR. These estimates decrease slightly with increasing censoring proportion. The estimates for early censoring are slightly less biased than for random or late censoring at the higher censoring proportions. The bias is not heavily influenced by sample size.

Crossing Hazards

Survival times are generated from the Weibull distribution where $\gamma=2.5$, $\alpha=0.3$ for group 1, and $\gamma=0.9$, $\alpha=2$ for group 2. The AHR is 15.4 for this situation. The percent of the bias for the mean Cox model estimate relative to the AHR is given in Table 4.

The bias of the Cox estimates tends to be much smaller for 10% and 25% censoring proportions compared to the 50% censoring proportion. For 50% censoring, the Cox model tends to overestimate the AHR. The bias decreases with increasing sample size, especially for high censoring proportions.

Diverging Hazards

Survival times are generated from the Weibull distribution where $\gamma=0.9$, $\alpha=1.0$ for group 1, and $\gamma=1.5$, $\alpha=2$ for group 2. The AHR is 0.536 for this situation. The percent of the bias for the mean Cox model estimate relative to the AHR is given in Table 5.

The Cox estimates are larger for random and late censoring than for early censoring at the highest censoring proportion. Generally, the sample size has little effect on the bias. For early censoring, the percent bias is approximately 20% and is not strongly affected by sample size or censoring proportion.

Table 1. Proportional Hazards: percent bias of Cox model estimates relative to average hazard rate of 2.0.

Censoring	% Censored	Sample Size per Group		
		30	50	100
Random	10%	5.5	4.0	2.0
	25%	8.0	4.5	2.0
	50%	11.0	5.5	3.0
Early	10%	7.0	5.0	2.5
	25%	10.0	6.0	3.5
	50%	19.5	11.0	7.0
Late	10%	5.5	4.0	2.0
	25%	7.0	4.0	2.5
	50%	10.5	6.5	3.5

Table 2. Increasing Hazards: percent bias of Cox model estimates relative to average hazard rate of 1.20.

Censoring	% Censored	Sample Size per Group		
		30	50	100
Random	10%	- 6.7	- 7.5	- 8.3
	25%	- 9.2	-10.8	-10.8
	50%	-15.0	-17.5	-18.3
Early	10%	- 4.2	- 5.8	- 6.7
	25%	- 4.2	- 5.8	- 5.8
	50%	- 1.7	- 5.0	- 5.8
Late	10%	- 7.5	- 9.2	-10.0
	25%	-12.5	-14.2	-15.0
	50%	-20.8	-22.5	-23.3

Table 3. Decreasing Hazards: percent bias of Cox model estimates relative to average hazard rate of 0.441.

Censoring	% Censored	Sample Size per Group		
		30	50	100
Random	10%	- 2.0	- 3.2	- 3.2
	25%	- 4.3	- 5.7	- 5.9
	50%	- 9.5	-11.3	-12.2
Early	10%	- 1.4	- 2.5	- 2.3
	25%	- 2.7	- 3.6	- 3.6
	50%	- 5.4	- 5.9	- 6.6
Late	10%	- 2.0	- 3.4	- 3.6
	25%	- 4.9	- 6.8	- 7.3
	50%	-10.9	-12.9	-13.8

Table 4. Crossing Hazards: percent bias of Cox model estimates relative to average hazard rate of 15.4.

Censoring	% Censored	Sample Size per Group		
		30	50	100
Random	10%	5.8	- 7.1	-14.9
	25%	19.5	4.5	- 5.2
	50%	73.3	52.6	34.4
Early	10%	1.3	-11.0	-18.8
	25%	9.1	- 5.2	-15.6
	50%	32.5	8.4	- 6.5
Late	10%	- 1.9	-12.9	-19.5
	25%	- 0.6	- 5.8	- 8.4
	50%	100.6	81.8	67.5

Table 5. Diverging Hazards: percent bias of Cox model estimates relative to average hazard rate of 0.536.

Censoring	% Censored	Sample Size per Group		
		30	50	100
Random	10%	-16.2	-18.3	-19.2
	25%	-10.4	-12.9	-14.2
	50%	7.8	3.7	1.1
Early	10%	-19.0	-20.9	-22.0
	25%	-19.0	-21.3	-22.6
	50%	-18.8	-21.8	-23.7
Late	10%	-16.4	-18.5	-19.4
	25%	- 6.9	- 9.3	-10.4
	50%	18.5	13.9	12.3

Table 6. Converging Hazards: percent bias of Cox model estimates relative to average hazard rate of 7.15.

Censoring	% Censored	Sample Size per Group		
		30	50	100
Random	10%	- 8.9	-11.2	-12.2
	25%	- 5.6	- 8.3	- 9.4
	50%	4.0	1.9	- 0.6
Early	10%	- 9.4	-11.3	-12.4
	25%	- 6.2	- 8.8	-10.2
	50%	2.4	- 0.8	- 4.3
Late	10%	-10.2	-12.4	-13.1
	25%	- 7.3	- 8.4	- 8.1
	50%	10.5	9.2	8.1

Converging Hazards

Survival times are generated from the Weibull distribution where $\gamma=0.9$, $\alpha=6.0$ for group 1, and $\gamma=1.2$, $\alpha=1$ for group 2. The AHR is 7.15 for this situation. The percent of the bias for the median Cox model estimate relative to the AHR is given in Table 6. The median Cox estimate increases with increasing censoring proportion. The bias is not heavily influenced by sample size.

Conclusion

Just as with the classical maximum likelihood estimator, the maximum partial likelihood estimator is not unbiased, but it is asymptotically unbiased (Kotz & Johnson, 1985, p. 591-593). This behavior is evident in Table 1, where the Cox estimates can be seen to be larger than the AHR, but the bias decreases with increasing sample size regardless of the type of censoring or the censoring rate.

Table 7 shows those instances where the average percent bias exceeds 20% in absolute value; the entries are the percent bias averaged over sample size.

There is no serious bias for the proportional hazards case regardless of type of censoring or censoring rate. Similarly, there is no serious bias in the cases of decreasing or converging hazards.

Under-estimation occurs for increasing hazards at the 50% censoring rate with late censoring. It also occurs for diverging hazards with early censoring regardless of censoring rate. Over-estimation occurs for crossing hazards at the 50% censoring rate with random and late censoring.

One might suspect that late censoring would render the least biased estimates since such a data structure contains more information than early or random censoring. However, late censoring leads to severe bias for increasing and crossing hazards when the censoring proportion is high. For lower censoring proportions (25% or lower), there is no severe bias for any of the nonproportionality models except diverging hazards.

As a practical matter, one can obtain descriptive statistics from a given data set, including percent censored, sample sizes, and a

plot of the hazard curves. From this information, one can approximate the magnitude and nature of the risk of biased estimation of the hazard ratio by the Cox model. Generally, the least biased estimates are obtained for the lower censoring proportions (10% and 25%) except for diverging hazards. In terms of bias, early censoring is problematic only for diverging hazards; late censoring is problematic for increasing and crossing hazards with the 50% censoring rate; and random censoring is problematic for crossing hazards with the 50% censoring rate. The case corresponding to the least occurrence of severe bias is the one involving random censoring with a censoring rate of 25% or less.

In practice, the experimenter typically has some control over sample size and perhaps the censoring proportion. For instance, the experimenter may be able to minimize censoring proportion, depending on the situation, through effective study design and experimental protocol. Minimizing the censoring rate is generally recommended, especially for increasing and crossing hazards. Early censoring is appreciably affected by censoring proportion only for constant and crossing hazards. Sample size has the strongest effect on constant and crossing hazards, especially at higher censoring proportions, where higher sample sizes lead to less biased estimates.

In practical applications, the proportional hazards assumption is never met precisely. If the deviation from the proportional hazards assumption is severe, then remedial measures should be taken. However, in many instances the model diagnostics reveal only a small to moderate deviation from the proportional hazards assumption. In these cases, the Cox model estimate of the hazard ratio is used for interpretation purposes in the presence of small to moderate assumption violations. This study characterizes the consequences of this interpretation in terms of bias, taking into account censoring rate, type of censoring, type of nonproportional hazards, and sample size. The general results indicate that the percent bias relative to AHR is under 20% in all but a few specific instances, as outlined above.

Table 7. Percent bias of the average Cox regression model estimates of the hazard ratio relative to the AHR averaged over sample size.

Hazards	% censoring	Censoring		
		random	early	late
constant	10	*	*	*
	25	*	*	*
	50	*	*	*
increasing	10	*	*	*
	25	*	*	*
	50	*	*	-22
decreasing	10	*	*	*
	25	*	*	*
	50	*	*	*
crossing	10	*	*	*
	25	*	*	*
	50	53	*	83
diverging	10	*	-21	*
	25	*	-21	*
	50	*	-21	*
converging	10	*	*	*
	25	*	*	*
	50	*	*	*

*under 20% in absolute value

References

- Andersen, P. K., Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10, 1100-1120.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, 34, 187-220.
- Cox, D. R., (1975). Partial likelihood. *Biometrika*, 62, 269-276.
- Hosmer, D. W., Lemeshow, S. (1999). Regression modeling of time to event data. New York: John Wiley & Sons.
- Kalbfleisch, J. D., Prentice, R. L. (1981). Estimation of the average hazard ratio. *Biometrika*, 68, 105-112.
- Kaplan, E. L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457-481.
- Klein, J. P., Moeschberger, M. L. (1997). *Survival analysis: Techniques for censored and truncated data*. Springer: New York.
- Kotz, S., & Johnson, N. L., (1985). *Encyclopedia of Statistical Sciences*, Volume 6, New York: John Wiley & Sons.
- Lee, E. T. (1992). Statistical methods for survival data analysis, (2nd Ed). Oklahoma City: John Wiley & Sons.
- Ng'andu, N. H. (1997). An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model. *Statistics in Medicine*, 16, 611- 626.
- Persson, I. (2002). Essays on the assumption of proportional hazards in Cox regression. Acta Universitatis Upsaliensis. Unpublished Ph.D. dissertation.
- Quantin, C., Moreau, T., Asselain, B., Maccario, J., & Lellouch, J. A. (1996). Regression survival model for testing the proportional hazards hypothesis. *Biometrics*, 52, 874-885.
- Song, H. H. & Lee, S. (2000). Comparison of goodness of fit tests for the Cox proportional hazards model. *Communications in Statistics – Simulation and Computation*, 29, 187-206.
- Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics*, 9, 93-108.