The Annals of Statistics
1982, Vol. 10, No. 4, 1121–1124

COMMENTARY ON ANDERSEN AND GILL'S "COX'S REGRESSION MODEL FOR COUNTING PROCESSES: A LARGE SAMPLE STUDY"

By Steven G. Self and Ross L. Prentice[1]

The Johns Hopkins University; the University of Washington and the Fred Hutchinson Cancer Research Center

In this issue Andersen and Gill (hereafter AG) present a stimulating development of asymptotic distribution theory for the Cox regression model with time-dependent covariates. They use a counting process formulation for the failure time data and martingale covergence results. This approach involves such conditions as σ-algebra right continuity and predictable, locally bounded, covariate processes. In this commentary we consider the implications of such assumptions for likelihood factorization and covariate modeling. In particular, it is noted that the partial likelihood function modeled by AG cannot, in general, involve covariate measurements at the random failure times. Some related work by the authors on a partial likelihood function that may involve covariate values at the random failure times is briefly discussed. Assumptions under which the intensity process modeled by AG has a standard "hazard" function interpretation are described and some generalizations of the AG results are mentioned.

1. Introduction. The Cox (1972) regression model is now widely used in failure time studies, particularly in biomedical applications. Such a model along with the partial likelihood (Cox, 1975) method of estimation merely requires the hazard ratio (relative risk) to be some non-negative parametric (usually exponential) form as a function of regression variables, without any restriction on the "baseline" hazard function. Notationally a dependence of the hazard ratio on time is accommodated through the inclusion of time-dependent variables in the modeled regression vector. For example, the regression vector may include product terms between some basic regression variable and time. Of equal importance in terms of applications is the use of a Cox-type model when the basic regression variables are stochastic processes over time. For example, the regression variable may include cumulative exposure levels to some carcinogen in an epidemiologic study or may include a sequence of daily leukocyte counts in a clinical trail involving the use of immunosuppressive drugs. In such settings Cox models provide a powerful means for studying complex interrelationships and failure time mechanism (see Kalbfleisch and Prentice, 1980, and Oakes, 1981, for elaboration of these points).

The work of AG provides a rigorous development of Cox model asymptotic distribution theory that is general enough to allow locally bounded stochastic time-dependent covariates and to include certain classes of multivariate failure time problems. They use a counting process representation of the failure times and decompose the counting process into the sum of its cumulative intensity process (or compensator) and a local square integrable martingale. A Cox-type regression model is specified for the intensity processes and some general results on the asymptotic behaviour of stochastic integrals with respect to martingales yield the desired convergence results. Here we consider likelihood factorizations that will lead to the likelihood function maximized by AG and consider conditions under which the intensity process will have a standard hazard function interpretation. Particular attention is paid to the technical assumptions made by AG in relation to covariate specification and modeling. We also discuss a related partial likelihood function that involves covariate values at the random failure times.

Received April 1982; revised May 1982.

[1] This work was supported in part by NIH grants GM-28314 and GM-24472.

2. Partial likelihood and the intensity process. Let us begin by considering the overall likelihood function using the same probabilistic structure and notation (supplemented as necessary) as do AG. This structure involves an increasing class of σ-algebras $\{\mathscr{F}_t;\ t \in [0, 1]\}$ that includes failure time, censoring and covariate histories. In the AG notation $N = (N_1, \cdots, N_n)$ denotes a multivariate counting process (and hence each N_i has right continuous sample paths) such that N_i counts "failures" on the ith subject over the time period [0, 1]; in fact, N_i count only "observable" failures in that N_i can jump only when the ith subject is under observation (at times t such that $Y_i(t) = 1$). The indicator process Y_i appears as a factor in the intensity process λ_i modeled by AG. In order that the intensity process be predictable, AG require the sample paths of Y_i to be left continuous. The multivariate indicator process $Y = (Y_1, \cdots, Y_n)$ will be referred to as the censoring process. As new notation, let $X = (X_1, \cdots, X_n)$ denote basic covariate processes such that $X_i(t) = \{X_{i1}(t), X_{i2}(t), \cdots\}$ involves measurements taken on the ith subject at time t. The theorems applied by AG require the family of σ-algebras to be right continuous (i.e., $\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}_s$) so that $\mathscr{F}_t$ should be defined to include counting process information up to and including time t, and censoring process information up to and including time t^+ (in view of the left continuity of sample paths of Y this will be necessary if $\mathscr{F}_t$ is to include information on $Y(t)$ values and be right continuous). The sample paths of X should be taken to be right continuous (with left-hand limits) in order that $\mathscr{F}_t$ may include information on $X(t)$ values, but not on $X(t + \varepsilon)$ values for any $\varepsilon > 0$ since such values may not even conceptually exist (i.e., covariate processes may be randomly stopped by the corresponding failure times).

The overall likelihood function can now be written as a product integral from 0 to 1 of

$$P[\mathscr{F}_{(t^- + dt)} \mid \mathscr{F}_{t^-}] = \lim_{s \uparrow t} P[\mathscr{F}_{(s+dt)} \mid \mathscr{F}_s]$$

where $\mathscr{F}_{t-} = \bigcup_{s<t} \mathscr{F}_s$ includes all information on counting and covariate processes up to, but not including, time t along with all information on the censoring process up to and including time t. Questions of interest usually involve the dependence of the failure rate at t on preceding (in time) covariate and counting process histories. The likelihood function can be factored to isolate this dependence by setting

$$P[\mathscr{F}_{(t^- + dt)} \mid \mathscr{F}_{t^-}] = P[N(t^- + dt) \mid \mathscr{F}_{t^-}]P[\mathscr{F}_{(t^- + dt)} \mid \mathscr{F}_{t^-}, N(t^- + dt)] \tag{2.1}$$

where, for example,

$$P[N(t^- + dt) \mid \mathscr{F}_{t^-}] = \lim_{s \uparrow t} P[N(s + dt) \mid \mathscr{F}_s].$$

The first factor in (2.1) can be written as a product $\tilde{\lambda}_i(t)\, dt$ for any subject (at most one) failing at t and $1 - \tilde{\lambda}_i(t)\, dt$ for each subject not failing at t, where

$$\tilde{\lambda}_i(t) = \lim_{s \uparrow t} \lim_{h \downarrow 0} h^{-1} P[N_i(s + h) - N_i(s) = 1 \mid \mathscr{F}_s]. \tag{2.2}$$

Expression (2.2) determines a stochastic process $\tilde{\lambda} = (\tilde{\lambda}_1, \cdots, \tilde{\lambda}_n)$ that is precisely the intensity function modeled by AG, under some regularity. Specifically

$$\tilde{\lambda}_i(t^+) = \lim_{h \downarrow 0} h^{-1} P[N_i(t + h) - N_i(t) = 1 \mid \mathscr{F}_t] \tag{2.3}$$

which equals the corresponding intensity process value $\lambda_i(t^+)$ if, for example, each λ_i is bounded by an integrable random variable (Aalen, 1978).

At a time t at which a failure occurs $P[N(t^- + dt) \mid \mathscr{F}_{t-}]$ can be further factorized as

$$\begin{aligned} P[N(t^- + dt) \mid \mathscr{F}_{t^-}] &= P[N(t^- + dt) \neq N(t^-) \mid \mathscr{F}_{t^-}] \\ &\cdot P[N_i(t^- + dt) - N_i(t^-) = 1 \mid \mathscr{F}_{t^-}, N(t^- + dt) \neq N(t^-)] \end{aligned} \tag{2.4}$$

the second factor of which, from (2.2), can be written

$$\lambda_i(t) / \textstyle\sum_{\ell=1}^{n} \lambda_\ell(t), \tag{2.5}$$

under the regularity mentioned above. It is the product of terms (2.5) over distinct failure times that constitutes the function $C(\beta, 1)$ maximized by AG. The above development shows it to be a partial likelihood function. They consider such maximization after specifying

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp\{\beta_0' Z_i(t)\}, \tag{2.6}$$

where $Z_i = (Z_{i1}, \cdots, Z_{ip})$ is an adapted column vector of p regression processes ($i = 1, \cdots, n$). In order to apply martingale convergence results AG require the sample paths of Z_i to be left continuous with right-hand limits (and so to be predictable and locally bounded). Aside from these requirements it is evident from the above development that the sample paths of Z_i can be any data-analyst-defined function of $\mathscr{F}_{t-}$. Note, however, that $Z(t)$ may not, in general, include functions of the corresponding basic covariate value $X(t)$ since $\mathscr{F}_{t-}$ only involves information on $X(u)$ for $u < t$. Within the context in which AG work, the numerator of (2.5) therefore may not include covariate information at the (random) failure times in contrast to the usual perspective of the Cox model partial likelihood (eg. AG expression (1.2)).

AG do not comment on the role of standard assumptions concerning the independence of failure times on distinct study subjects and the independence of the censoring mechanism in their development. While such assumptions are not necessary for the partial likelihood development or for the corresponding asymptotic distribution theory, they will usually be necessary for the intensity processes to have a useful interpretation. Beginning with (2.3), an independent censoring mechanism may be defined as one requiring, for each $i = 1, \cdots, n$,

$$\tilde{\lambda}_i(t^+) = \lim_{h\downarrow 0} h^{-1}P[N_i(t+h) - N_i(t) = 1 \mid \{N_\ell(u); X_\ell(u); 0 \le u \le t\}, \ell = 1, \cdots, n]$$

at all times t at which the ith subject is under observation ($Y_i(t) = 1$). The addition of an independent failure time assumption between subjects would require

$$\tilde{\lambda}_i(t^+) = \lim_{h\downarrow 0} h^{-1}P[N_i(t+h) - N_i(t) = 1 \mid \{N_i(u), X_i(u); 0 \le u \le t\}]$$

at all t values such that $Y_i(t) = 1$. The right side of this expression has an ordinary "hazard" function interpretation so that the parameters involved in the modeling of $\lambda = (\lambda_1, \cdots, \lambda_n)$ will have a clear interpretation under independent failure time and independent censoring assumptions and under sufficient regularity to ensure $\tilde{\lambda} = \lambda$.

3. An alternate partial likelihood function. The overall likelihood function can alternatively be written as a product integral from 0 to 1 of $P[\mathscr{F}_{(t+dt)} \mid \mathscr{F}_t]$. A factorization like (2.1) (with t^- everywhere replaced by t) may be written whence it can be noted that $P[N(t+dt) \mid \mathscr{F}_t]$ is a product over $i = 1, \cdots, n$ of $h_i(t)dt$ if subject i fails at t or $1 - h_i(t)dt$ if subject i does not fail at t, where

$$h_i(t) = \lim_{h\downarrow 0} h^{-1}P[N_i(t+h) - N_i(t) = 1 \mid \mathscr{F}_t]. \tag{3.1}$$

Note that $h_i(t) = \tilde{\lambda}_i(t^+)$. A further factorization like (2.4) (again with t in place of t^-) then leads to a partial likelihood function which is a product of terms

$$h_i(t)/\textstyle\sum_{\ell=1}^n h_\ell(t), \tag{3.2}$$

over each failure time. Note that $h_1(t) = (h_1(t), \cdots, h_n(t))$ may be modeled in terms of available information on covariate and counting processes up to and including time t (or the censoring process up to t^+) since such information is included in $\mathscr{F}_t$. For example, the Cox model (2.6) for $\tilde{\lambda}_i(t)$ gives

$$h_i(t) = \tilde{\lambda}_i(t^+) = Y_i(t^+)\lambda_0(t^+)\exp\{\beta_0' Z_i(t^+)\}. \tag{3.3}$$

Under such a model the partial likelihood function from (3.2) addresses a different set of questions concerning the dependence of failure rate on covariate histories than does the partial likelihood from (2.4), since the failure rate $h(t)$ conditions on $X(t)$ (and thereby on jumps $X(t) - X(t^-)$ in the covariate process). Of course, the two partial likelihood functions will coincide if it is assumed that counting and covariate processes cannot jump simultaneously. Otherwise it is evident that some regularity in the covariate processes at the failure times will need to be imposed in order that the maximum partial likelihood function $\hat{\beta}$ from (3.2) be well behaved. As a rather extreme special case suppose the covariate processes jump so that some component of $X(T)$ is identically equal to a fixed value x_0 at each (random) failure time T. The likelihood factorization and partial likelihood development described in this section then completely degenerates.

To demonstrate the consistency of $\hat{\beta}$ from (3.2), conditions similar to those of AG may be used together with a generalization of Tsiatis's (1981) argument (see Self, 1981; Self and Prentice, 1982). It is, however, necessary also to introduce a smoothness condition on the covariate path just prior to failure. A sufficient condition is

$$E[Z(t) \mid N(t) \neq N(t^-)] = E[Z(t^+) \mid N(t) \neq N(t^-)].$$

By using a generalization of Tsiatis's (1981) expansion of the score statistic, the asymptotic normality of $\hat{\beta}$ may also be demonstrated but at the expense of introducing stronger moment conditions than do AG as well as additional smoothness (e.g., bounded variation) and regularity conditions on the covariate paths. The absence of such conditions in the AG development points to the power of the martingale convergence results that they use. It also seems evident that the partial likelihood function from (2.5), rather than that from (3.2), should generally be used in applications.

4. Generalizations. There have been a number of other generalizations of the original Cox (1972) model that are of practical importance. The methods employed by AG seem well suited to most of these. For example, the baseline hazard function may be permitted to differ among each of a fixed number of strata (as mentioned by AG). Time-dependent strata may be accommodated by allowing $Y_i(t)$ to indicate the stratum assignment as well as the "observation" status for the ith subject at time t. The martingale results will also apply to some problems with "forward going" semi-Markov processes (Voelkel and Crowley, 1982) and would appear to generalize directly to the estimation of type-specific intensity functions in competing risk problems. We will present elsewhere a generalization of the AG results to Cox-type models in which the exponential form of hazard function dependence on covariates is replaced by an arbitrary (non-negative) parametric form.

REFERENCES

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist* **10** 1100–1120

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B* **34** 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

OAKES, D. (1981). Survival times: Aspects of partial likelihood. *Internat. Statist. Reviews* 235–264.

SELF, S. (1981). Proportional hazards model with time-dependent covariates: Large sample theory and related topics. Ph.D. Thesis, Department of Biostatist., Univ. of Washington.

SELF, S. and PRENTICE, R. L. (1982). Cox's regression model with time-dependent covariates: Large sample theory. Technical Report, Dept. of Biostatist., Univ. of Washington.

TSIATIS, A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9** 93–108.

VOELKEL, J. and CROWLEY, J. (1982). Nonparametric inference for a class of semi-Markov processes with censored observations. Technical Report #15, Department of Statist., Univ. of Wisconsin.

DEPARTMENT OF BIOSTATISTICS
SCHOOL OF HYGIENE AND PUBLIC HEALTH
JOHNS HOPKINS UNIVERSITY
615 N. WOLFE STREET
BALTIMORE, MARYLAND 21205

PROGRAM IN EPIDEMIOLOGY AND BIOSTATISTICS
THE FRED HUTCHINSON CANCER RESEARCH CENTER
1124 COLUMBIA STREET
SEATTLE, WASHINGTON 98104