APPENDIX A: PROOFS OF CONSISTENCY AND ASYMPTOTIC NORMALITY

The conditions for the consistency and asymptotic normality of $\hat{\beta}$ and $\hat{\Lambda}_0(\tau)$ in the Cox model were given in Andersen and Gill (1982), which used martingales to simplify and generalize the asymptotic results of Cox (1975) and Tsiatis (1981). Conditions for the more general relative risk model were given in Prentice and Self (1983). Here, we outline the most important of these conditions and point out their implications for the use of relative risk regression models in infectious disease epidemiology.

A.1 Regularity conditions

Assume all observations take place at infectiousness ages in $[0, \mathcal{T}]$ for some finite $\mathcal{T}$. Let $m = Y(0^+) = \lim_{\tau \downarrow 0} Y(\tau)$ be the number of pairs ij that were at risk of infectious contact from i to j while under observation. Let n_m denote the number of individuals that constitute the m pairs. Define the following functions (Prentice and Self, 1983):

$$S_m^{(0)}(\beta, \tau) = \frac{1}{m} Y(\beta, \tau) = \sum_{j=1}^{n_m} \sum_{i \neq j} r\big(\beta^\mathsf{T} X_{ij}(\tau)\big) Y_{ij}(\tau),$$

$$S_m^{(1)}(\beta, \tau) = \frac{\partial}{\partial \beta} S_m^{(0)}(\tau) = \frac{1}{m} \sum_{j=1}^{n_m} X_{ij}(\tau) r'\big(\beta^\mathsf{T} X_{ij}(\tau)\big) Y_{ij}(\tau), \text{ and}$$

$$S_m^{(2)}(\beta, \tau) = \frac{1}{m} \sum_{j=1}^{n_m} \sum_{i \neq j} X_{ij}(\tau)^{\otimes 2} \Big((\ln r)'\big(\beta^\mathsf{T} X_{ij}(\tau)\big) \Big)^2 r\big(\beta^\mathsf{T} X_{ij}(\tau)\big) Y_{ij}(\tau).$$

Note that $S_m^{(0)}$ is real-valued, $S_m^{(1)}$ is $b \times 1$ vector-valued, and $S_m^{(2)}$ is $b \times b$ matrix-valued. Now let

$$E_m(\beta, \tau) = \frac{S_m^{(1)}(\beta, \tau)}{S_m^{(0)}(\beta, \tau)} \text{ and} \tag{A.1}$$

$$V_m(\beta, \tau) = \frac{S_m^{(2)}(\beta, \tau)}{S_m^{(0)}(\beta, \tau)} - E_m(\beta, \tau)^{\otimes 2} \tag{A.2}$$

be the values of $E(\beta, \tau)$ and $V(\beta, \tau)$, respectively, based on observations of m pairs at risk of transmission.

For consistency and asymptotic normality of $\sqrt{m}(\hat{\beta} - \beta_0)$, we have the following sufficient conditions (Andersen and Gill, 1982; Prentice and Self, 1983):

A. (Finite interval) $\Lambda_0(\mathcal{T}) < \infty$.

B. (Regression function positivity) There exists a neighborhood $\mathcal{B}_0$ of β_0 such that $r\big(\beta^\mathsf{T} X_{ij}(\tau)\big)$ is locally bounded away from zero for all ij and all $\beta \in \mathcal{B}_0$.

C. (Asymptotic stability) There exists a neighborhood $\mathcal{B} \subseteq \mathcal{B}_0$ of β_0 and functions $s^{(0)}, s^{(1)}, s^{(2)}$ defined on $\mathcal{B} \times [0, \mathcal{T}]$ such that

$$\sup_{\beta \in \mathcal{B}, \tau \in [0, \mathcal{T}]} \| S_m^{(k)}(\beta, \tau) - s^{(k)}(\beta, \tau) \| \xrightarrow{\text{P}} 0 \text{ as } m \to \infty \tag{A.3}$$

for $k = 0, 1, 2$. Here, $\|x\|$ is $|x|$ for real x, $\max\big(|x_1|, \ldots, |x_b|\big)$ for vector x, and $\max\big(|x_{11}|, \ldots, |x_{bb}|\big)$ for matrix x. Asymptotic properties of the Cox model depend only on convergence of these three functions. For more general relative risk functions, convergence of four additional functions is also required (Prentice and Self, 1983).

D. (Asymptotic regularity) The functions $s^{(0)}(\beta, \tau), \ldots, s^{(2)}(\beta, \tau)$ are bounded on $\mathcal{B} \times [0, \mathcal{T}]$ and continuous in β uniformly in τ. In addition, $s^{(0)}$ is bounded away from zero and has first and second derivatives with respect to β on $\mathcal{B} \times [0, \mathcal{T}]$. Finally, let $e(\beta, \tau) = \frac{s^{(1)}(\beta, \tau)}{s^{(0)}(\beta, \tau)}$ and $v(\beta, \tau) = \frac{s^{(2)}(\beta, \tau)}{s^{(0)}(\beta, \tau)} - e(\beta, \tau)^{\otimes 2}$. Then

$$\Sigma = \int_0^{\mathcal{T}} v(\beta_0, u) s^{(0)}(\beta_0, u) \lambda_0(u) \, \mathrm{d}u \tag{A.4}$$

is positive definite.

E. (Asymptotic stability of the observed information matrix)

$$\sup_{\beta \in \mathcal{B}} \int_0^{\mathcal{T}} \frac{1}{m^2} \sum_{j=1}^{n_m} \sum_{i \neq j} \|X_{ij}(u)\|^4 \Big((\ln r)''\big(\beta^\top X_{ij}(u)\big) \Big)^2 r\big(\beta_0^\top X_{ij}(u)\big) \lambda_0(u) \, \mathrm{d}u \xrightarrow{\text{P}} 0. \tag{A.5}$$

F. (Lindeberg condition)

$$\frac{1}{\sqrt{m}} \sup_{\tau, ij} \Big\| X_{ij}(\tau) (\ln r)'\big(\beta_0^\top X_{ij}(\tau)\big) \Big\| \xrightarrow{\text{P}} 0, \tag{A.6}$$

where the supremum is over all $\tau \in [0, \mathcal{T}]$ and all ij such that $Y_{ij}(\tau) = 1$.

Condition F is automatically fulfilled if the covariates X_{ij} are bounded. In the Cox model, conditions B and E are automatically fulfilled because $\exp(x) > 0$ and $(\ln r)''(x) = 0$ for all real x. With only slight modification, these conditions also guarantee consistency and asymptotic normality in a stratified relative-risk regression model (Andersen and Borgan, 1985).

For the methods in this paper, the most important constraint is that $s^{(0)}(\beta, \tau)$ is bounded away from zero. This has two implications for infectious disease data that have no counterpart in standard survival data. First, the infectious period must be $\geq \mathcal{T}$ with positive probability. Second, the hazard of infection in the susceptible j from a randomly chosen ij at risk of transmission must have a finite mean as $m \to \infty$. To state this requirement more precisely, let

$$Y_{\cdot j}(\tau) = \sum_{i \neq j} Y_{ij}(\tau) \Rightarrow \sum_{j=1}^{n_m} Y_{\cdot j}(0^+) = m. \tag{A.7}$$

Now let $D_{ij} = Y_{\cdot j}(0^+)Y_{ij}(0^+)$ be the number of infectors to which j is exposed if ij was at risk of transmission and $D_{ij} = 0$ otherwise. If we randomly choose a pair ij at risk of transmission and look at the number of infectors to which j is exposed, its expected value is

$$\frac{1}{m}\sum_{j=1}^{n_m}\sum_{i\neq j} D_{ij} = \frac{1}{m}\sum_{j=1}^{n_m} Y_{\cdot j}(0^+)^2. \tag{A.8}$$

For $s^{(0)}(\beta,\tau)$ to be bounded away from zero, we must have

$$\limsup_{m\to\infty} \frac{1}{m}\sum_{j=1}^{n_m} Y_{\cdot j}(0^+)^2 < \infty. \tag{A.9}$$

If not, a randomly chosen ij has a j with a hazard of infection that becomes infinite as $m \to \infty$, which implies $\mathbb{E}[\,Y_{ij}(\tau)\,] \to 0$ and $s^{(0)}(\beta,\tau) \to 0$ for all $\tau > 0$. In practice, this constraint implies that large-sample distributions are useful when both the number of pairs m and the number of susceptibles are large such that the largest value of $Y_{\cdot j}(0^+) \ll m$.

There is no similar constraint on the number of susceptibles exposed to each infectious person. In theory, we could have m susceptibles exposed to a single infectious person without violating the regularity conditions (as long as his or her infectious period was $\geq \mathcal{T}$). This is because the contact intervals in all pairs ij for a fixed i are assumed to be independent of each other and independent of the infectious period of i conditional on the covariate processes $X_{ij}(\tau)$.

A.2 Asymptotic properties of $U(\beta_0,\tau)$, $\hat{\beta}$, and $\hat{\Lambda}_0(\tau)$

Let $U_m(\beta_0,\tau)$ denote the score process based on observations of m pairs ij at risk of transmission when who-infects-whom is observed, let $\hat{\beta}_m$ denote the corresponding maximum partial likelihood estimate, and let $\hat{\Lambda}_{0,m}(\tau)$ denote the corresponding Breslow estimate of the baseline hazard. Under the conditions of the last section, we have the following results as $m \to \infty$ (Andersen and Gill, 1982; Prentice and Self, 1983):

1. Asymptotic normality of the score: $\frac{1}{\sqrt{m}}U(\beta_0,\mathcal{T}) \xrightarrow{\text{D}} N\big(0,\Sigma\big)$.
2. Consistency of $I(\beta_0)$ and $\mathcal{I}(\beta_0)$: $\frac{1}{m}I(\beta_0) \xrightarrow{\text{P}} \Sigma$ and $\frac{1}{m}\mathcal{I}(\beta_0) \xrightarrow{\text{P}} \Sigma$.
3. Consistency of $\hat{\beta}$: $\hat{\beta}_m \xrightarrow{\text{P}} \beta_0$.
4. Asymptotic normality of $\hat{\beta}$: $\sqrt{m}(\hat{\beta}-\beta_0) \xrightarrow{\text{D}} N\big(0,\Sigma^{-1}\big)$.
5. Consistency of $I(\hat{\beta})$ and $\mathcal{I}(\hat{\beta})$: $\frac{1}{m}I(\hat{\beta}) \xrightarrow{\text{P}} \Sigma$ and $\frac{1}{m}\mathcal{I}(\hat{\beta}) \xrightarrow{\text{P}} \Sigma$.
6. Convergence of $\sqrt{m}\big(\hat{\Lambda}_0(\tau) - \Lambda_0(\tau)\big)$ to a mean-zero Gaussian process with independent increments.

7. Asymptotic independence of

$$\left(\frac{\partial}{\partial\beta}\hat{\Lambda}(\beta^*,\tau)\right)^{\mathsf{T}}\sqrt{m}(\hat{\beta}-\beta_0)$$

and

$$m\int_0^\tau \frac{1}{Y(\beta_0,u)^2}\,\mathrm{d}N(u).$$

8. Continuity of $\frac{\partial}{\partial\beta}\hat{\Lambda}(\beta^*,\tau)$: $\frac{\partial}{\partial\beta}\hat{\Lambda}(\beta_m,\tau)\xrightarrow{\text{P}}\frac{\partial}{\partial\beta}\hat{\Lambda}(\beta_0,\tau)$ if $\beta_m\xrightarrow{\text{P}}\beta$.

APPENDIX B: ASYMPTOTIC VARIANCE OF BASELINE HAZARD ESTIMATES

Andersen and Gill (1982) showed that $\sqrt{m}\big(\hat{\Lambda}_0(\tau)-\Lambda_0(\tau)\big)$ converges to a mean-zero Gaussian martingale in the Cox model for standard survival data, and this result was extended to more general relative risk functions by Prentice and Self (1983). Under the conditions given in Appendix , these derivations extend directly to infectious disease data.

B.1 Who-infects-whom is observed

Expanding $\hat{\Lambda}_0(\tau)-\Lambda_0(\tau)$ gives us

$$\begin{aligned}\sqrt{m}\Big(\hat{\Lambda}_0(\tau)-\Lambda_0(\tau)\Big) &= \sqrt{m}\Big(\hat{\Lambda}(\hat{\beta},\tau)-\hat{\Lambda}(\beta_0,\tau)\Big)\\ &\quad+\sqrt{m}\Big(\hat{\Lambda}(\beta_0,\tau)-\Lambda_0^*(\tau)\Big)\\ &\quad+\sqrt{m}\big(\Lambda_0^*(\tau)-\Lambda_0(\tau)\big),\end{aligned} \tag{B.1}$$

where $\Lambda_0^*(\tau)=\int_0^\tau \mathbf{1}_{Y(u)>0}\lambda_0(u)\,\mathrm{d}u$. By a first-order Taylor expansion, the first term in (B.1) is

$$\left(\frac{\partial}{\partial\beta}\hat{\Lambda}(\beta^*,\tau)\right)^{\mathsf{T}}\sqrt{m}\big(\hat{\beta}-\beta_0\big) \tag{B.2}$$

for some β^* on the line segment between β_0 and $\hat{\beta}$. Using the Doob-Meyer decomposition, the second term in (B.1) can be written

$$\sqrt{m}\int_0^\tau \frac{\mathbf{1}_{Y(u)>0}}{Y(\beta_0,u)}\,\mathrm{d}M(u), \tag{B.3}$$

which is a martingale with the optional variation process

$$m\int_0^\tau \frac{1}{Y(\beta_0,u)^2}\,\mathrm{d}N(u). \tag{B.4}$$

The third term in (B.1) is zero For all τ such that $Y(\tau) > 0$. Under the regularity conditions of Appendix , the first and second terms are asymptotically independent, so the asymptotic variance of (B.1) is

$$\left(\frac{\partial}{\partial\beta}\hat{\Lambda}(\hat{\beta},\tau)\right)^{\mathsf{T}}\left(\frac{1}{m}I(\hat{\beta})\right)^{-1}\left(\frac{\partial}{\partial\beta}\hat{\Lambda}(\hat{\beta},\tau)\right)+\int_0^\tau \frac{m}{Y(\hat{\beta},u)^2}\,\mathrm{d}N(u) \tag{B.5}$$

for all τ such that $Y(\tau) > 0$.

B.2 Who-infects-whom is not observed

By an expansion similar to that in equation (B.1),

$$\begin{aligned}\sqrt{m}\Big(\widetilde{\Lambda}_0(\tau)-\Lambda_0(\tau)\Big) &= \sqrt{m}\Big(\widetilde{\Lambda}_{\widetilde{\beta},\widetilde{\lambda}}(\widetilde{\beta},\tau)-\widetilde{\Lambda}_{\widetilde{\beta},\widetilde{\lambda}}(\beta_0,\tau)\Big)\\ &\quad+\sqrt{m}\Big(\widetilde{\Lambda}_{\widetilde{\beta},\widetilde{\lambda}}(\beta_0,\tau)-\widetilde{\Lambda}_{\beta_0,\lambda_0}(\beta_0,\tau)\Big)\\ &\quad+\sqrt{m}\Big(\widetilde{\Lambda}_{\beta_0,\lambda_0}(\beta_0,\tau)-\Lambda_0^*(\tau)\Big)\\ &\quad+\sqrt{m}\big(\Lambda_0^*(\tau)-\Lambda_0(\tau)\big).\end{aligned} \tag{B.6}$$

The fourth term in (B.6) is zero for all τ at which $Y(\tau) > 0$.

By a first-order Taylor expansion, the first term in (B.6) equals

$$\left(\frac{\partial}{\partial\beta}\widetilde{\Lambda}_{\widetilde{\beta},\widetilde{\lambda}}(\beta^*,\tau)\right)^{\mathsf{T}}\sqrt{m}\big(\widetilde{\beta}-\beta_0\big) \tag{B.7}$$

for some β^* on the line segment between β_0 and $\widetilde{\beta}$, where

$$\frac{\partial}{\partial\beta}\widetilde{\Lambda}_{\widetilde{\beta},\widetilde{\lambda}}(\beta,\tau) = -\int_0^\tau \frac{\frac{\partial}{\partial\beta}Y(\beta,u)}{Y(\beta,u)^2}\,\mathrm{d}\widetilde{N}(u|\widetilde{\beta},\widetilde{\lambda}). \tag{B.8}$$

Its contribution to the variance is

$$\left(\frac{\partial}{\partial\beta}\widetilde{\Lambda}_{\widetilde{\beta},\widetilde{\lambda}}(\beta_0,\tau)\right)^{\mathsf{T}}\left(\frac{1}{m}\widetilde{I}(\beta_0)\right)^{-1}\left(\frac{\partial}{\partial\beta}\widetilde{\Lambda}_{\widetilde{\beta},\widetilde{\lambda}}(\beta_0,\tau)\right). \tag{B.9}$$

The second term in (B.6) can be rewritten

$$\sqrt{m}\int_0^\tau \frac{1}{Y(\beta_0,u)}\Big(\mathrm{d}\widetilde{N}(u|\widetilde{\beta},\widetilde{\lambda})-\mathrm{d}\widetilde{N}(u|\beta_0,\lambda_0)\Big) \tag{B.10}$$

For each j, we have $\int_0^\infty \mathrm{d}\widetilde{N}(u|\beta,\lambda) = 1$ if j was infected and 0 otherwise. Thus, the term in parentheses is the sum a subset of the random variables $\delta_{ij} = p_{ij}(\widetilde{\beta},\widetilde{\lambda}) - p_{ij}(\beta_0,\lambda_0)$, which have sum zero for each j. Since the δ_{ij} are asymptotically independent for different j and $Y(\beta_0,u) = O_P(m)$, the integral behaves asymptotically like a mean of independent random variables with mean zero and variance $O(\widetilde{\beta} - \beta_0)$. Therefore, the second term of (B.6) is $O_P(\widetilde{\beta} - \beta_0)$ and converges in probability to zero as $m \to \infty$.

The third term in (B.6) can be evaluated using the conditional variance formula. The expression inside the parentheses has the variance

$$\mathbb{E}_{\beta_0,\lambda_0}\big[\,\hat{\sigma}^2_{\mathbf{v}}(\beta_0,\tau)\,\big] + \mathrm{Var}_{\beta_0,\lambda_0}\big(\hat{\Lambda}_{\mathbf{v}}(\beta_0,\tau)\big) = \\ \int_0^\tau \frac{1}{Y(\beta_0,u)^2}\,\mathrm{d}\widetilde{N}(u|\beta_0,\lambda_0) + \mathbb{E}_{\beta_0,\lambda_0}\big[\,\hat{\Lambda}_{\mathbf{v}}(\beta_0,\tau)^2\,\big] - \widetilde{\Lambda}_{\beta_0,\lambda_0}(\beta_0,\tau)^2, \quad \text{(B.11)}$$

where

$$\hat{\sigma}^2_{\mathbf{v}}(\beta,\tau) = \int_0^\tau \frac{1}{Y(\beta,u)^2}\,\mathrm{d}N(u|\mathbf{v}). \quad \text{(B.12)}$$

Since each infected person has only one infector and infectors can be chosen independently given the observed data,

$$\mathbb{E}_{\beta_0,\lambda_0}\big[\,\hat{\Lambda}_{\mathbf{v}}(\beta_0,\tau)^2\,\big] = \widetilde{\Lambda}_{\beta_0,\lambda_0}(\widetilde{\beta},\tau)^2 - \sum_{j=1}^n\bigg(\int_0^\tau \frac{1}{Y(\beta_0,u)}\,\mathrm{d}\widetilde{N}_{\cdot j}(u|\beta_0,\lambda_0)\bigg)^2 \\ + \int_0^\tau \frac{1}{Y(\beta_0,u)^2}\,\mathrm{d}\widetilde{N}(u|\beta_0,\lambda_0), \quad \text{(B.13)}$$

where $\widetilde{N}_{\cdot j}(u|\beta,\lambda) = \sum_{i\neq j}\widetilde{N}_{ij}(u|\beta,\lambda)$. Therefore, the total variance contribution of the third term in (B.6) reduces to

$$2\int_0^\tau \frac{m}{Y(\beta_0,u)^2}\,\mathrm{d}\widetilde{N}(u|\beta_0,\lambda_0) - \sum_{j=1}^n\bigg(\int_0^\tau \frac{\sqrt{m}}{Y(\beta_0,u)}\,\mathrm{d}\widetilde{N}_{\cdot j}(u|\beta_0,\lambda_0)\bigg)^2. \quad \text{(B.14)}$$

Since only the first and third terms of (B.6) are asymptotically nonzero, all that remains is to look at their covariance. Let $N_{ij}(\tau|\mathbf{v})$ denote the value of $N_{ij}(\tau)$ that we would have calculated had we observed the transmission network $\mathbf{v}$. Then the corresponding value of the score $U(\beta,\tau)$ is

$$U_{\mathbf{v}}(\beta,\tau) = \sum_{j=1}^n\sum_{i\neq j}\int_0^\tau \frac{\partial}{\partial\beta}\ln\frac{r\big(\beta^{\mathsf{T}}X_{ij}(u)\big)}{Y(\beta,u)}\,\mathrm{d}N(u|\mathbf{v}) \quad \text{(B.15)}$$

and the corresponding covariance of $U(\beta,\tau)$ and $\hat{\Lambda}(\beta,\tau)$ is

$$\kappa_{\mathbf{v}}(\beta,\tau) = \sum_{j=1}^n\sum_{i\neq j}\int_0^\tau \frac{1}{Y(\beta,u)}\bigg(\frac{\partial}{\partial\beta}\ln\frac{r\big(\beta^{\mathsf{T}}X_{ij}(u)\big)}{Y(\beta,u)}\bigg)\,\mathrm{d}N_{ij}(u|\mathbf{v}). \quad \text{(B.16)}$$

By the conditional covariance formula,

$$\mathrm{Cov}\Big(\widetilde{U}_{\beta_0,\lambda_0}(\beta_0,\lambda_0),\widetilde{\Lambda}_{\beta_0,\lambda_0}(\beta_0,\tau)\Big) = \mathrm{Cov}_{\beta_0,\lambda_0}\big(U_{\mathbf{v}}(\beta_0,\tau),\hat{\Lambda}_{\mathbf{v}}(\beta_0,\tau)\big) \\ + \mathbb{E}_{\beta_0,\lambda_0}\big[\,\kappa_{\mathbf{v}}(\beta_0,\tau)\,\big] \quad \text{(B.17)}$$

By an argument similar to that leading to (B.14), this reduces to

$$2\sum_{j=1}^{n}\sum_{i\neq j}\int_0^{\tau}\frac{1}{Y(\beta_0,u)}\left(\frac{\partial}{\partial\beta}\ln\frac{r\big(\beta_0^{\mathsf{T}}X_{ij}(u)\big)}{Y(\beta_0,u)}\right)\mathrm{d}\widetilde{N}_{ij}(u|\beta_0,\lambda_0)$$
$$-\sum_{j=1}^{n}\left(\int_0^{\tau}\frac{1}{Y(\beta_0,u)}\,\mathrm{d}\widetilde{N}_{\cdot j}(u|\beta_0,\lambda_0)\right)U_{\cdot j}(\beta_0,\tau). \tag{B.18}$$

In the limit of large m, both terms in (B.18) act like means of random variables with mean zero and finite variance, so they converge in probability to zero. Since $\widetilde{\beta}$ is a function of the expected score, this implies that the first and third terms of equation (B.6) are asymptotically independent.

Combining all of these results, the asymptotic variance of (B.6) is

$$\left(\frac{\partial}{\partial\beta}\widetilde{\Lambda}_{\widetilde{\beta},\widetilde{\lambda}}(\beta_0,\tau)\right)^{\mathsf{T}}\left(\frac{1}{m}\widetilde{I}(\beta_0)\right)^{-1}\left(\frac{\partial}{\partial\beta}\widetilde{\Lambda}_{\widetilde{\beta},\widetilde{\lambda}}(\beta_0,\tau)\right)$$
$$+2\int_0^{\tau}\frac{m}{Y(\beta_0,u)^2}\,\mathrm{d}\widetilde{N}(u|\beta_0,\lambda_0)-\sum_{j=1}^{n}\left(\int_0^{\tau}\frac{\sqrt{m}}{Y(\beta_0,u)}\,\mathrm{d}\widetilde{N}_{\cdot j}(u|\beta_0,\lambda_0)\right)^2. \tag{B.19}$$

References

Per Kragh Andersen and Ørnulf Borgan. Counting process models for life history data: A review. *Scandinavian Journal of Statistics*, 12:97–158, 1985.

Per Kragh Andersen and Richard D. Gill. Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10:1100–1120, 1982.

David R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.

Ross L. Prentice and Steven G. Self. Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Annals of Statistics*, 11:804–813, 1983.

Anastasios A. Tsiatis. A large-sample study of Cox's regression model. *Annals of Statistics*, 9:93–108, 1981.