Web Supplementary Materials for "Identifying subjects benefiting from additional information for better prediction"

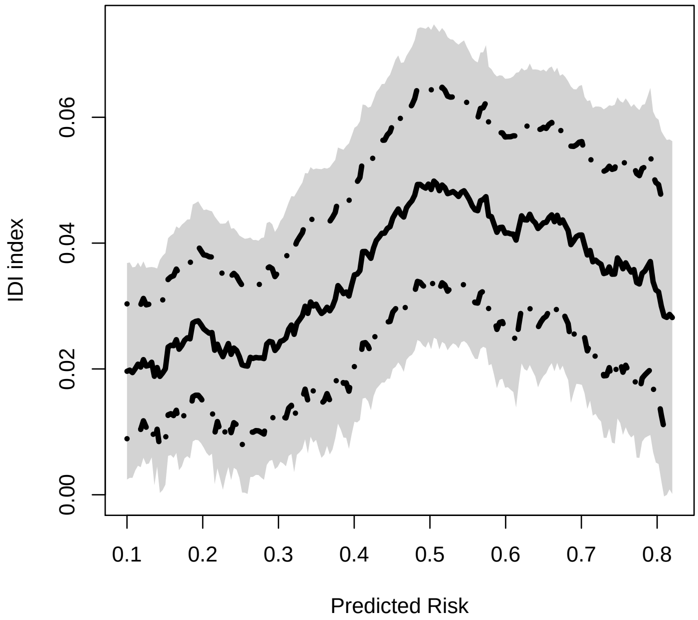
L. Tian

Department of Preventive Medicine, Northwestern University, Chicago, IL 60611.

T. Cai and L.J. Wei

Department of Biostatistics, Harvard University, Boston, MA 02115.

Figure 1 Point estimate for IDI index and the corresponding 0.95 point-wise (dashed lines) and simultaneous (shaded regions) confidence intervals for the screened population of the TRACE study.



Appendix

Appendix A *Large sample properties of* $\widehat{\mathcal{D}}_1(\cdot)$, $\widehat{\mathcal{D}}_2(\cdot)$ *and* $\widehat{\Delta}(\cdot)$

To justify the asymptotic properties of the proposed estimators, certain *smooth* regularity conditions are needed for the distance function $D(\cdot,\cdot)$ and its corresponding predictor. Here, we consider the case that the distance function is $D(Y,\widehat{Y}) = |Y-\widehat{Y}|$ for continuous and $w_0^{(1-Y)} w_1^Y |Y-\widehat{Y}|$ for binary responses, where w_0 and w_1 are given positive numbers. Furthermore, when Y is continuous, we let $\widehat{Y}_1(\beta' x) = g_1(\beta' x)$ and $\widehat{Y}_2(\theta' w) = g_2(\theta' w)$, and when Y is binary, let $\widehat{Y}_1(\beta' x) = I\{g_1(\beta' x) \geqslant c\}$ and $\widehat{Y}_2(\theta' w) = I\{g_2(\theta' w) \geqslant c\}$ for some constant c. Similar arguments can be used to justify other cases.

Suppose that β_0 and θ_0 are interior points of their compact parameter spaces. We assume that the random vector X and W are bounded above and thus $\sup_{\boldsymbol{\beta}} |\widehat{Y}_1(\beta' X)| + \sup_{\boldsymbol{\theta}} |\widehat{Y}_2(\theta' W)|$ is bounded by a constant $\mathcal{Y}_0$. Let Ω denote the set of z such that J_z is properly contained in the support of $\beta_0' X$. First, we show that the above estimators are uniformly consistent over Ω. To this end, with a slight abuse of notation, we let

$$\widetilde{\mathcal{D}}_1(z,\beta) = \frac{\sum_{i=1}^n D\{Y_i, \widehat{Y}_1(\beta' X_i)\} I\{g_1(\beta' X_i) \in J_z\}}{\sum_{i=1}^n I\{g_1(\beta' X_i) \in J_z\}},$$

$$\widetilde{\mathcal{D}}_2(z,\beta,\theta) = \frac{\sum_{i=1}^n D\{Y_i, \widehat{Y}_2(\theta' W_i)\} I\{g_1(\beta' X_i) \in J_z\}}{\sum_{i=1}^n I\{g_1(\beta' X_i) \in J_z\}},$$

$\mathcal{D}_1(z,\beta) = E[D\{Y, \widehat{Y}_1(\beta' X)\} | g_1(\beta' X) \in J_z]$ and $\mathcal{D}_2(z,\beta,\theta) = E[D\{Y, \widehat{Y}_2(\theta' W)\} | g_1(\beta' X) \in J_z]$. Note that $\mathcal{D}_1(z) = \mathcal{D}_1(z,\beta_0)$ and $\mathcal{D}_2(z) = \mathcal{D}_2(z,\beta_0,\theta_0)$. When Y is binary, it is not difficult to see that the processes $I\{g_1(\beta' X_i) \in J_z\} = I\{\beta' X_i \in g_1^{-1}(J_z)\}$, $D\{Y_i, \widehat{Y}_1(\beta' X_i)\} = w_1 Y_i I(\beta' X_i \geqslant c) + w_0(1-Y_i) I(\beta' X_i < c)$ and $D\{Y_i, \widehat{Y}_2(\theta' W_i)\} = w_1 Y_i I(\theta' W_i \geqslant c) + w_0(1-Y_i) I(\theta' W_i < c)$ have finite pseudo dimensions. Therefore, by the stability of the manageability of processes, $D\{Y_i, \widehat{Y}_1(\beta' X_i)\}\, I\{g_1(\beta' X_i) \in J_z\}$ and $D\{Y_i, \widehat{Y}_2(\theta' W_i)\} I\{g_1(\beta' X_i) \in J_z\}$ are manageable. Furthermore, it is not difficult to see that $D\{Y_i, \widehat{Y}_1(\beta' X_i)\} I\{g_1(\beta' X_i) \in J_z\}$ and $D\{Y_i, \widehat{Y}_2(\theta' W_i)\} I\{g_1(\beta' X_i) \in J_z\}$ are bounded by an envelop function $|Y_i| + \mathcal{Y}_0$ up to multiplying a universal constant. It then follows from the uniform law of large numbers (Pollard, 1990, Ch. 8) that $\sup_{z,\beta} |\widetilde{\mathcal{D}}_1(z,\beta) - \mathcal{D}_1(z,\beta)| + \sup_{z,\beta,\theta} |\widetilde{\mathcal{D}}_2(z,\beta,\theta) - \mathcal{D}_2(z,\beta,\theta)|$ converges to 0, in probability, where the sup is taken over Ω and the compact parameter spaces of β and θ. When Y is continuous, $D\{Y_i, \widehat{Y}_1(\beta' X_i)\} = |Y_i - g_1(\beta' X_i)|$ and $D\{Y_i, \widehat{Y}_2(\theta' W_i)\} = |Y_i - g(\theta' W_i)|$ are Lipshitz continuous in β and θ, respectively. Therefore, the classes of functions $\{D\{y, \widehat{Y}_1(\beta' x)\} \mid \beta\}$ and $\{D\{y, \widehat{Y}_2(\theta' w)\} \mid \theta\}$ are Donsker (van der Vaart and Wellner, 1996, Theorem 2.7.11). Following similar arguments, the uniform law of large number can be applied to show the convergence of $\sup_{z,\beta} |\widetilde{\mathcal{D}}_1(z,\beta) - \mathcal{D}_1(z,\beta)|$ and $\sup_{z,\beta,\theta} |\widetilde{\mathcal{D}}_2(z,\beta,\theta) - \mathcal{D}_2(z,\beta,\theta)|$. This uniform convergence, together with the convergence property of $\widehat{\beta}$ and $\widehat{\theta}$, implies the uniform consistency of $\widehat{\mathcal{D}}_1(z) = \widetilde{\mathcal{D}}_1(z,\widehat{\beta})$ and $\widehat{\mathcal{D}}_2(z) = \widetilde{\mathcal{D}}_2(z,\widehat{\beta},\widehat{\theta})$. The consistency of $\widehat{\Delta}(\cdot)$ follows accordingly.

Next, we show that the processes $\widehat{\mathcal{D}}_1(\cdot)$, $\widehat{\mathcal{D}}_2(\cdot)$ and $\widehat{\Delta}(\cdot)$ after standardization are asymptotically normal. First, let $T = (Y, U', V')'$. It follows from Appendix 1 of Tian et. al. (2007),

$$n^{\frac{1}{2}} \begin{pmatrix} \widehat{\beta} - \beta_0 \\ \widehat{\theta} - \theta_0 \end{pmatrix} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \begin{pmatrix} \psi_1(T_i) \\ \psi_2(T_i) \end{pmatrix} + o_p(1),$$

where

$$\psi_1(T) = [E\{\dot{g}_1(\beta_0'X)XX'\}]^{-1}X\{Y - g_1(\beta_0'X)\} \text{ and } \psi_2(T) = [E\{\dot{g}_2(\theta_0'W)WW'\}]^{-1}W\{Y - g_2(\theta_0'W)\}.$$

Now, let $\widehat{\mathcal{W}}_1(z,\beta) = n^{\frac{1}{2}}\{\widetilde{\mathcal{D}}_1(z,\beta) - \mathcal{D}_1(z,\beta)\}$, $\widehat{\mathcal{W}}_2(z,\beta,\theta) = n^{\frac{1}{2}}\{\widetilde{\mathcal{D}}_2(z,\beta,\theta) - \mathcal{D}_2(z,\beta,\theta)\}$,

$$\xi_1(z,\beta,T_i) = \frac{I\{g_1(\beta'X_i) \in J_z\}\left[D\{Y_i, \widehat{Y}_1(\beta'X_i)\} - \mathcal{D}_1(z,\beta)\right]}{\text{pr}\{g_1(\beta'X) \in J_z\}},$$

and

$$\xi_2(z,\beta,\theta,T_i) = \frac{I\{g_1(\beta'X_i) \in J_z\}\left[D\{Y_i, \widehat{Y}_2(\theta'W_i)\} - \mathcal{D}_2(z,\beta,\theta)\right]}{\text{pr}\{g_1(\beta'X) \in J_z\}}.$$

Writing

$$\begin{aligned} &\widehat{\mathcal{W}}_1(z,\beta) - n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_1(z,\beta,T_i) \\ = &- \{\widetilde{\mathcal{D}}_1(z,\beta) - \mathcal{D}_1(z,\beta)\} n^{-\frac{1}{2}} \sum_{i=1}^{n} \left[\frac{I\{g_1(\beta'X_i) \in J_z\} - \text{pr}\{g_1(\beta'X) \in J_z\}}{\text{pr}\{g_1(\beta'X) \in J_z\}}\right] \end{aligned}$$

and using a Slutsky Theorem for random process, one can show that $\sup_{z,\beta} \left|\widehat{\mathcal{W}}_1(z,\beta) - n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_1(z,\beta,T_i)\right| \to 0$ in probability as $n \to \infty$ (van der Vaart and Wellner, 1996, Example 1.4.7). Similarly, we have $\sup_{z,\beta,\theta} \left|\widehat{\mathcal{W}}_2(z,\beta,\theta) - n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_2(z,\beta,\theta,T_i)\right| = o_p(1)$. Furthermore, using the similar argument used to show consistency of $\widehat{\mathcal{D}}_k(z), k = 1, 2$, it can be shown that the classes of functions $\{\xi_1(z,\beta,t) \mid z,\beta\}$ and $\{\xi_2(z,\beta,\theta,t) \mid z,\beta,\theta\}$ are Donsker for either continuous or binary Y. Therefore, the processes $\widehat{\mathcal{W}}_1(z,\beta)$ and $\widehat{\mathcal{W}}_2(z,\beta,\theta)$ are tight.

This, together with the above asymptotic expansion of $n^{\frac{1}{2}}(\widehat{\beta}-\beta_0)$ and $n^{\frac{1}{2}}(\widehat{\theta}-\theta_0)$, implies that

$$
\begin{aligned}
\widehat{\mathcal{W}_1}(z) &= \widehat{\mathcal{W}_1}(z,\widehat{\beta}) + n^{\frac{1}{2}}\left\{\mathcal{D}_1(z,\widehat{\beta}) - \mathcal{D}_1(z,\beta_0)\right\} \\
&= \widehat{\mathcal{W}_1}(z,\beta_0) + \left\{ \frac{\partial \mathcal{D}_1(z,\beta)}{\partial \beta}\Big|_{\beta_0} \right\}' n^{\frac{1}{2}}(\widehat{\beta}-\beta_0) + o_p(1) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} \left[\xi_1(z,\beta_0,T_i) + \left\{ \frac{\partial \mathcal{D}_1(z,\beta)}{\partial \beta}\Big|_{\beta_0} \right\}' \psi_1(T_i) \right] + o_p(1) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} \eta_1(z,\beta_0,T_i) + o_p(1) \\
\widehat{\mathcal{W}_2}(z) &= \widehat{\mathcal{W}_2}(z,\widehat{\beta},\widehat{\theta}) + n^{\frac{1}{2}}\left\{\mathcal{D}_2(z,\widehat{\beta},\widehat{\theta}) - \mathcal{D}_2(z,\beta_0,\theta_0)\right\} \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} \left[\xi_2(z,\beta_0,\theta_0,T_i) + \left\{ \frac{\mathcal{D}_2(z,\beta,\theta)}{\partial \beta}\Big|_{(\beta_0,\theta_0)} \right\}' \psi_1(T_i) + \left\{ \frac{\partial \mathcal{D}_2(z,\beta,\theta)}{\partial \theta}\Big|_{(\beta_0,\theta_0)} \right\}' \psi_2(T_i) \right] + o_p(1) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} \eta_2(z,\beta_0,\theta_0,T_i) + o_p(1),
\end{aligned}
$$

where $o_p(1)$ is uniform in z. Firstly, one can show that the classes of functions $\{\eta_1(z,\beta_0,t) \mid z\}$ and $\{\eta_2(z,\beta_0,\theta_0,t) \mid z\}$ are Donsker for either continuous or binary Y. Since $\text{pr}\{g_1(\beta_0'X) \in J_z\}$, $D_1(z,\beta_0)$, $D_2(z,\beta_0,\theta_0)$, $\partial\mathcal{D}_1(z,\beta)/\partial\beta\big|_{\beta_0}$, $\mathcal{D}_2(z,\beta,\theta)/\partial\beta\big|_{(\beta_0,\theta_0)}$ and $\partial\mathcal{D}_2(z,\beta,\theta)/\partial\theta\big|_{(\beta_0,\theta_0)}$ are continuous in z, all these functions are bounded by a constant for $z \in \Omega$. Furthermore, $\text{pr}\{g_1(\beta_0'X) \in J_z\}$ is bounded away from 0 for $z \in \Omega$. Therefore, $|\eta_1(z,\beta_0,T_i)| \leqslant C_0\{|Y_i| + \mathbf{1}'|\psi_1(T_i)|\}$ and $|\eta_2(z,\beta_0,\theta_0,T_i)| \leqslant C_0\{|Y_i| + \mathbf{1}'|\psi_2(T_i)|\}$, where $\mathbf{1}$ is a column vector with all elements being 1 and C_0 is a generic constant. It then follows from a functional central limit theorem (Pollard, 1990, Ch. 10) that the processes $\widehat{\mathcal{W}_1}(\cdot)$ and $\widehat{\mathcal{W}_2}(\cdot)$ converge weakly to zero-mean Gaussian processes. The weak convergence of $\widehat{\mathcal{W}}(\cdot)$ follows accordingly.

To approximate the distribution of the processes $\widehat{\mathcal{W}_1}(\cdot)$, $\widehat{\mathcal{W}_2}(\cdot)$ and $\widehat{\mathcal{W}}(\cdot)$, we consider their respective perturbed versions

$$
\begin{aligned}
\mathcal{W}_1^*(z) &= n^{\frac{1}{2}} \left\{ \frac{\sum_{i=1}^n [D\{Y_i,\widehat{Y}_1(\widehat{\beta}'X_i)\} - \widehat{\mathcal{D}}_1(z)] I\{g(\widehat{\beta}'X_i) \in J_z\} G_i}{\sum_{i=1}^n I\{g(\widehat{\beta}'X_i) \in J_z\}} + \widetilde{\mathcal{D}}(z,\widehat{\theta}^*) - \widehat{\mathcal{D}}_1(z) \right\}, \\
\mathcal{W}_2^*(z) &= n^{\frac{1}{2}} \left\{ \frac{\sum_{i=1}^n [D\{Y_i,\widehat{Y}_2(\widehat{\theta}'W_i)\} - \widehat{\mathcal{D}}_2(z)] I\{g(\widehat{\beta}'X_i) \in J_z\} G_i}{\sum_{i=1}^n I\{g(\widehat{\beta}'X_i) \in J_z\}} + \widetilde{\mathcal{D}}(z,\widehat{\beta}^*,\widehat{\theta}^*) - \widehat{\mathcal{D}}_2(z) \right\},
\end{aligned}
$$

and $\mathcal{W}^*(z) = \mathcal{W}_1^*(z) - \mathcal{W}_2^*(z)$, where

$$
\begin{aligned}
\widehat{\beta}^* &= \widehat{\beta} + \left\{ \sum_{i=1}^n \dot{g}_1(\widehat{\beta}'X_i) X_i X_i' \right\}^{-1} \sum_{i=1}^n X_i \left\{ Y_i - g_1(\widehat{\beta}'X_i) \right\} G_i \\
\widehat{\theta}^* &= \widehat{\theta} + \left\{ \sum_{i=1}^n \dot{g}_2(\widehat{\theta}'W_i) W_i W_i' \right\}^{-1} \sum_{i=1}^n W_i \left\{ Y_i - g_2(\widehat{\theta}'W_i) \right\} G_i,
\end{aligned}
$$

and $\{G_1,...,G_n\}$ are independent standard normal random variables that are independent of the data. It follows from the same arguments as given above and similar arguments as in Appendix 4 of Cai et. al. (2005) that the limiting distributions of $\mathcal{W}_1^*(\cdot)$, $\mathcal{W}_2^*(\cdot)$ and $\mathcal{W}^*(\cdot)$, conditional on the data, are the same as those of $\widehat{\mathcal{W}}_1(\cdot)$, $\widehat{\mathcal{W}}_2(\cdot)$ and $\widehat{\mathcal{W}}(\cdot)$, respectively, on Ω. Since $\mathrm{pr}(\widehat{\Omega} \subset \Omega) \to 1$, the confidence interval given in (2.10) of the paper is asymptotically valid for any $z \in \widehat{\Omega}$. Furthermore, noting the fact that $\sup_{\widehat{\Omega}} |\mathcal{W}_l^*(z)/\sigma_{\mathcal{W}_l^*(z)}|$ and $\sup_{\widehat{\Omega}} |\widehat{\mathcal{W}}_l(z)/\sigma_{\widehat{\mathcal{W}}_l(z)}|$ are asymptotically equivalent to $\sup_{\Omega_{d_1,d_2}} |\mathcal{W}_l^*(z)/\sigma_{\mathcal{W}_l^*(z)}|$ and $\sup_{\Omega_{d_1,d_2}} |\widehat{\mathcal{W}}_l(z)/\sigma_{\widehat{\mathcal{W}}_l(z)}|$, respectively, where $\Omega_{d_1,d_2} \subset \Omega$ is the limit of $\widehat{\Omega}$, the asymptotical confidence band over the random region $\widehat{\Omega}$ given in (2.11) of the paper is valid as well. Similarly, one may justify the validity of the pointwise and simultaneous confidence intervals given in (2.12) and (2.13) in the paper by noting that $\widetilde{\Omega}_{d_3}$, the limit of $\widetilde{\Omega}$, is a subset of Ω and $\sigma_{\mathcal{W}(z)}$ is uniformly bounded below by a positive constant for $z \in \widetilde{\Omega}$.

Appendix B *Large sample properties of crossvalidated estimators*

For each partition $\mathcal{I}_k$, let $\widehat{\beta}_{(-k)}$ and $\widehat{\theta}_{(-k)}$ be the estimated β_0 and θ_0 using data not in $\mathcal{I}_k$ via (2.2) and (2.5), respectively,

$$\widehat{\mathcal{D}}_{1k}(z,\beta) = \frac{\sum_{i\in\mathcal{I}_k} D\{Y_i, \widehat{Y}_1(\beta' X_i)\} I\{g_1(\beta' X_i) \in J_z\}}{\sum_{i\in\mathcal{I}_k} I\{g_1(\beta' X_i) \in J_z\}},$$

and

$$\widehat{\mathcal{D}}_{2k}(z,\beta,\theta) = \frac{\sum_{i\in\mathcal{I}_k} D\{Y_i, \widehat{Y}_2(\theta' W_i)\} I\{g_1(\beta' X_i) \in J_z\}}{\sum_{i\in\mathcal{I}_k} I\{g_1(\beta' X_i) \in J_z\}}.$$

Since K is small relative to n, $\widehat{\mathcal{D}}_{1k}(z) = \widehat{D}_{1k}(z, \widehat{\beta}_{(-k)})$ is consistent. Then, it follows from the same argument in Appendix A, $n^{\frac{1}{2}}\{\widehat{\mathcal{D}}_{1k}(z) - \mathcal{D}_1(z)\} = n^{\frac{1}{2}}\{\widehat{D}_{1k}(z, \widehat{\beta}_{(-k)}) - \mathcal{D}_1(z, \beta_0)\}$ is asymptotically equivalent to

$$n^{-\frac{1}{2}} K \sum_{i=1}^n I(\tau_i = k)\xi_1(z, \beta_0, T_i) + \left\{ \frac{\partial \mathcal{D}_1(z,\beta)}{\partial \beta} \Big|_{\beta_0} \right\}' n^{\frac{1}{2}}(\widehat{\beta}_{(-k)} - \beta_0),$$

where $\{\tau_i; i = 1, \cdots, n\}$ are n exchangeable discrete random variables uniformly distributed over $\{1, 2, \cdots, K\}$, independent of the data, and $\sum_{i=1}^n I(\tau_i = k) \approx n/K, k = 1, \cdots, K$. It follows from the same argument in Appendix 3 of Tian et. al. (2007) that conditional on the observed $\{\tau_i, i = 1, \cdots, n\}$

$$\widehat{\beta}_{(-k)} - \beta_0 = \frac{K}{n(K-1)} \sum_{i=1}^n I(\tau_i \neq k)\psi_1(T_i) + o_p(n^{-1/2}).$$

Then using the same argument in Appendix A, one can show that

$$\widetilde{\mathcal{W}}_1(z) = \frac{n^{\frac{1}{2}}}{K} \sum_{k=1}^K \left\{ \widehat{\mathcal{D}}_{1k}(z) - \mathcal{D}_1(z) \right\} = \frac{n^{-\frac{1}{2}}}{K} \sum_{i=1}^n \sum_{k=1}^K \left\{ I(\tau_i = k) K \xi_1(z, \beta_0, T_i) + \frac{K I(\tau_i \neq k)\psi_1(T_i)}{K-1} \right\} + o_p(1).$$

Since $\sum_{k=1}^{K} I(\tau_i = k) = 1$ and $\sum_{k=1}^{K} I(\tau_i \neq k) = K - 1$, it is straightforward to show that $\widetilde{\mathcal{W}}_1(z)$ is asymptotically equivalent to $\widehat{\mathcal{W}}_1(z)$ and thus the distribution of $\widetilde{\mathcal{W}}_1(\cdot)$ can be approximated by that of $\mathcal{W}_1^*(\cdot)$ conditional on the partition indicators $\{\tau_i, i = 1, \cdots, n\}$. Similar arguments can be used to show that the distributions of $\widetilde{\mathcal{W}}_2(\cdot)$ and $\widetilde{\mathcal{W}}(\cdot)$ can be approximated by those of $\mathcal{W}_2^*(\cdot)$ and $\mathcal{W}^*(\cdot)$, respectively.

References

Cai, T., Tian, L., and Wei, L. J. (2005). Semiparametric Box-cox power transformation models for cnesored survival observations. *Biometrika* **92**, 619–32.

Pollard, D. (1990). Empirical Processes: Theory and Applications. NSF-CMBS Regional Conference Series in Probability and Statistics 2. Hayward, CA: Institute of Mathematical Statistics.

Tian, L., Cai, T., Goetghebeur, E., and Wei, L. J. (2007). Model evaluation based on the distribution of estimated absolute prediction error. *Biometrika* **94**, 297–311.

van der Vaart, A. W. and Wellner, J. (1996). Weak convergence and empirical processes. Springer-Verlag, New York.