Inverse probability weighted estimation in survival analysis.

Andrea Rotnitzky[1] and James Robins[2]
[1]Department of Biostatistics, Harvard School of Public Health
[2]Departments of Biostatistics and Epidemiology, Harvard School of Public Health

1 Introduction

Modern epidemiologic and clinical studies aimed at analyzing a time to an event endpoint T routinely collect, in addition to (possibly censored) information on T, high dimensional data often in the form of baseline (i.e. time- independent covariates $V(0)$) and time-varying covariates $V(t), t > 0$, measured at frequent intervals. Scientific interest, however, often focuses on a low dimensional functional $\beta = \beta(F_X)$ of the distribution F_X of the (intended) full data $X = \left(T, \overline{V}(T)\right)$ where $\overline{V}(t) \equiv \{V(u) : 0 \leq u \leq t\}$. Inverse probability weighted augmented (AIPW) estimators of β meet the analytic challenge posed by these high dimensional data because they are consistent and asymptotically normal (CAN) under models that do not make assumptions about the parts of F_X that are of little scientific interest. As such, they are not subject to biases induced by misspecification of models for these secondary parts of F_X.

AIPW estimators were originally introduced by Robins and Rotnitzky, 1992, as part of a general estimating function methodology in coarsened, i.e. incompletely observed, data models under non- or semi-parametric models for arbitrary full data configurations X when the data are coarsened at random (CAR) (Heitjan and Rubin, 1991, Jacobsen and Keiding, 1995, Gill, van der Laan and Robins, 1997) and the coarsening, i.e. censoring or missingness, mechanism is either known or correctly modelled. The AIPW estimators generalized and made efficient the non-augmented inverse probability weighted estimators proposed by Koul, Susarla and van Ryzin, 1981, and Keiding, Holst and Green, 1989. Robins and Rotnitzky, 1992, derived their methodology drawing from the modern theory of semiparametric efficiency due to Bickel, Klaassen, Ritov and Wellner, 1993, Newey, 1990, van der Vaart, 1988, 1991, among others. In this article we restrict the discussion to coarsened data in the form of right censored failure time data. In this setting, the full data are $X = \left(T, \overline{V}\left(T\right)\right)$, the observed data are $Y = \left(\widetilde{T} = \min\left(T, C\right), \Delta = I\left(T \leq C\right), \overline{V}\left(\widetilde{T}\right)\right)$ where C is a censoring variable, and C and T are continuos positive random variables. Furthermore, the CAR assumption is equivalent to

$$\lambda_C\left(u|X\right) = \lambda_C\left(u|\overline{V}\left(u\right)\right) \text{ for all } u \geq 0 \tag{1}$$

where $\lambda_C\left(u|\cdot\right) = \lim_{h\to 0^+}\Pr\left(u \leq C < u+h|\cdot, C \geq u, T \geq u\right)$ is the cause specific hazard for censoring. Therefore, the coarsening mechanism is determined by the stochastic process $G \equiv G\left(\cdot\right)$ where $G\left(u\right) \equiv \exp\left\{-\int_0^u \lambda_C\left(t|\overline{V}\left(t\right)\right)\ dt\right\}$.

2 AIPW estimators under CAR

2.1 Motivation: the curse of dimensionality

When the data are CAR, the likelihood $\mathcal{L}_n\left(F_X, G\right)$ based on n i.i.d. copies of Y factorizes as $\mathcal{L}_n\left(F_X, G\right) = \mathcal{L}_{1,n}\left(F_X\right)\mathcal{L}_{2,n}\left(G\right)$ where G denotes the coarsening mechanism, i.e. the conditional distribution of the observed data Y given the full data X, and F_X is the cumulative distribution function of X. Thus, for models in which G and F_X are variation independent, any method that obeys the likelihood principle must result in the same inference about β regardless of whether G is known, completely unknown or known to follow a model. However, under non- or big semi-parametric models for F_X, Robins and Ritov, 1997, have shown that due to the curse of dimensionality, with high dimensional coarsened at random data any method of inference that obeys the likelihood principle and thus ignores G must perform poorly in realistic sample sizes as the following example illustrates.

Example: Suppose that no covariates $V\left(t\right)$ are measured for any $t > 0$ and to simplify the notation let W denote $V\left(0\right)$. Under CAR, $\mathcal{L}_n\left(F_X, G\right) = \mathcal{L}_{1,n}\left(F_X\right)\mathcal{L}_{2,n}\left(G\right)$ where $\mathcal{L}_{1,n}\left(F_X\right) = \prod_{i=1}^n f_{T|W}\left(\widetilde{T}_i|W_i\right)^{\Delta_i}\left(1 - F_{T|W}\left(\widetilde{T}_i|W_i\right)\right)^{1-\Delta_i} f_W\left(W_i\right)$. Suppose that we are interested in estimating $\beta = \Pr(T \le t)$ for a fixed t. Because $\beta = \beta(F_X) = E_{F_W}\left\{F_{T|W}(t|W)\right\}$, its MLE $\widehat{\beta}$ is equal to $\beta\left(\widehat{F}_X\right)$ where $\widehat{F}_X$ is the MLE of F_X. Since the MLE of F_W is the empirical c.d.f. of W, then $\widehat{\beta} = n^{-1}\sum_i \widehat{F}_{T|W}(t|W_i)$. The MLE $\widehat{F}_{T|W}$ of $F_{T|W=W_i}$ is given by the Kaplan-Meier estimator based on the subsample $\{Y_i : W = W_i\}$. If W is continuous each subsample consists of one observation, so the non-parametric MLE of $F_{T|W=W_i}$ assigns probability 1 to $\widetilde{T}_i$; however, when $\Delta_i = 0$ the NPMLE of $F_{T|W=W_i}$ assigns probability 0 to the interval $\left[0, \widetilde{T}_i\right]$ and it is undefined on $\left(\widetilde{T}_i, \infty\right)$. Thus, when W is continuous the NPMLE of $F_{T|W=W_i}$ is undefined for some observed values of W_i and hence the MLE of β is also undefined. One could assume that $F_{T|W}$ was smooth in W and use multivariate smoothing techniques. However when W is high dimensional, $F_{T|W}$ would not be well estimated with the moderate sample sizes found in practice, because no two units would have values of W close enough to allow the borrowing of information needed for smoothing. Thus, unrealistically large sample sizes would be required for any estimator of β to have an approximately centered normal sampling distribution with variance small enough to be of substantive use $\Diamond$

AIPW estimators depend on a model for G and thus violate the likelihood principle, yet they yield estimators that are well behaved with moderate sample sizes. Locally efficient AIPW estimators of β (defined in the next section) simultaneously correct for bias due to dependent censoring attributable to the covariate process $V(t)$ and recover information from the censored observations by non-parametrically exploiting the correlation between the process $V(t)$ observed up to the censoring time and the unobserved failure time T.

Even under CAR, because of the curse of dimensionality, well behaved estimators of β in finite samples do not exist unless one imposes additional restrictions on either the coarsening mechanism G or on the non- or semiparametric model for F_X. Hence the best that can be hoped for is an estimator that is CAN under the CAR assumption (1) when either (but not necessarily both) a lower dimensional model for G or a lower dimensional model for F_X is correct. Such estimator, when it exists, is called doubly-robust (DR). Scharfstein, Rotnitzky and Robins, 1999, Robins, 2000, Robins, Rotnitzky and van der Laan, 2000 and van der Laan and Robins (2003) provide a broad theory of double robustness in CAR models. Using this theory, these authors show that the locally efficient AIPW estimators (described in the next section) are doubly robust. Robins and Rotnitzky, 2001, provide a summary of known results on double-robust estimation, including estimation in non-ignorable models, i.e. when CAR is not assumed.

2.2 Locally efficient AIPW estimation

Suppose that $\beta = \beta(F_X)$ is a smooth $k \times 1$ parameter under a non or semiparametric model $\mathcal{M}_{Full}$ for F_X. That is, $\beta(F_X)$ is estimable at rate $\sqrt{n}$ under all laws F_X in model $\mathcal{M}_{Full}$ when X is fully observed for all n sample units. The general algorithm for the construction of locally efficient AIPW estimators of a $k \times 1$ vector $\beta(F_X)$ starts with the specification of a full-data *orthogonal* estimating function $D(\beta, \rho) = d(X; \beta, \rho)$ for β. This is a $k \times 1$ vector function of the full data X, of β and, possibly, of a nuisance parameter ρ, such that each component of $D(\beta(F_X), \rho(F_X))$ has mean zero and covariance zero with any nuisance score under F_X, for all F_X in $\mathcal{M}_{Full}$. We need to allow the orthogonal estimating equation to possibly depend on a nuisance parameter ρ so as to make the general methodology applicable to a broad class of estimation problems. For instance, in the Cox proportional hazards model with time independent covariates, $\rho \equiv \rho(\cdot)$ where $\rho(u)$ is the mean of the covariate among subjects who fail at time u.

The full-data estimating function $D(\beta, \rho)$ gives rise to the observed data estimating function

$$U\{D(\beta, \rho), G\} - A(h_{F_X}, G) \tag{2}$$

The term $U\{D(\beta, \rho), G\}$ is an inverse probability weighted estimating function defined as

$$U\{D(\beta, \rho); G\} = \frac{\tau D(\beta, \rho)}{G(T^*)}$$

where T^* is the minimum time such that $D(\beta, \rho)$ is observed and $\tau = I(T^* < C)$ is the indicator that $D(\beta, \rho)$ is observed. The critical point of inverse probability weighting is that when $\Pr(\tau = 1|X) > 0$, the estimating function is *unbiased inverse probability* estimating function because, under CAR, $\Pr(\tau = 1|X) = G(T^*)$ and, thus, $E_G[U\{D(\beta, \rho); G\}|X] = D(\beta, \rho)$.

The second term in (2) is a mean zero augmentation term which, for any function $h\left(u, \overline{V}(u)\right)$, is defined as

$$A(h,G) = \int_0^{\widetilde{T}} \left[h\left(u,\overline{V}(u)\right)/G(u)\right] dM_C(u)$$

where $dM_C(u) = I\left(\widetilde{T} = u, \Delta = 0\right) - I\left(\widetilde{T} \geq u\right)\lambda_C\left(u|\overline{V}(u)\right)du$. The function h_{F_X} in (2) depends on F_X and G and is defined as

$$h_{F_X}\left(u,\overline{V}(u)\right) = -E\left[UD(\beta,\rho)\,|\overline{V}(u), T \geq u\right]$$

A doubly robust locally efficient AIPW estimator $\widehat{\beta}(D)$ that uses $D(\beta,\rho)$ is the solution to

$$\sum_{i=1}^{n}\left[U\left\{D_i\left(\beta,\widehat{\rho}(\beta)\right);\widehat{G}\right\} - A_i\left(h_{\widehat{F}_X},\widehat{G}\right)\right] = 0 \tag{3}$$

where $\widehat{F}_X$ and $\widehat{G}$ are the maximum likelihood estimators of F_X and G under parametric or semiparametric working models $\mathcal{M}_{work} \subset \mathcal{M}_{full}$ for F_X and $\mathcal{G}_{work}$ for G and $\widehat{\rho}(\beta)$ is an estimator of ρ such that $\widehat{\rho}(\beta)$ evaluated at the true β converges at an appropriate rate (usually $n^{1/4}$) to $\rho(F_X)$ if either, but not necessarily both, working models are true. See van der Laan and Robins, 2002, for construction of $\widehat{\rho}(\beta)$. The estimator $\widehat{\beta}(D)$ has the following properties:

a) If $\Pr(\tau = 1|X) > \sigma > 0$, $\widehat{\beta}(D)$ is doubly robust. That is, provided $\widehat{F}_X$ and $\widehat{G}$ converge at a sufficiently fast rate to F_X and G under $\mathcal{M}_{work}$ and $\mathcal{G}_{work}$ respectively, $\widehat{\beta}(D)$ is CAN in the union model that assumes that $F_X \in \mathcal{M}_{full}$,CAR and either $F_X \in \mathcal{M}_{work}$ or $G \in \mathcal{G}_{work}$.

b) There exists $D_{opt}(\beta,\rho)$ such that $\widehat{\beta}(D_{opt})$ is locally semiparametric efficient in the union model of part a) at the intersection submodel where both $\mathcal{M}_{work}$ and $\mathcal{G}_{work}$ are correct. Robins and Rotnitzky, 1992, and van der Laan and Robins (2002) show how to derive D_{opt}.

The previous results can be derived from the results in Robins and Rotnitzky, 1992, which provide a general representation for the influence functions and the efficient score of estimators of smooth parameters $\beta = \beta(F_X)$ of non- or semi-parametric models $\mathcal{M}_{Full}$ for distributions F_X of arbitrary full data configurations X under parametric, semiparametric or non-parametric CAR models for the censoring or missingness mechanism.

Example (continued): When $\beta = P(T \leq t)$ for a fixed t, then $D(\beta,\rho) = I(T \leq t) - \beta$ does not depend on a nuisance parameter and, because the full data model is non-parametric

$D(\beta,\rho)$ is, up to a multiplicative constant, the unique (orthogonal) unbiased estimating function. In this setting we have $T^{*}=\min(T,t)$ and

$$h_{F_X}(u,W)=E\{I(T\leq t)-\beta|W,T\geq u\}=F_{T|W,T\geq u}(t|W,T\geq u)-\beta$$

To obtain a locally efficient DR estimator of β, we specify a low dimensional, e.g. parametric, model $F_{T|W}(u|W;\eta)$ for $F_{T|W}(u|W)$ and compute the maximum likelihood estimator $\widehat{\eta}$ of η under the model. We leave the marginal distribution of W unrestricted so its MLE is the empirical distribution of W. In addition, we specify a low dimensional model for $G(u)$, for example, we may postulate a Cox proportional hazards model $\lambda_C(u|W)=\lambda_0(u)\exp(\gamma' m(W))$ and estimate γ with the Cox partial likelihood estimator $\widehat{\gamma}$ and $\lambda_0(u)$ with the Cox baseline hazard estimator $\widehat{\lambda}_0(u)$ regarding the censoring times as the outcomes and the failure times T as the censoring times for C. We then compute $\widehat{G}(u)=\prod_{\left\{i:\widetilde{T}_i\leq u,\Delta_i=0\right\}}\left[1-\widehat{\lambda}_0\left(\widetilde{T}_i\right)\exp\left(\widehat{\gamma}' m(W)\right)\right]$. We compute $h_{\widehat{F}_X,\widehat{G}}(u,W)$ using $\widehat{F}_{T|W}(u|W;\eta)=F_{T|W}(u|W;\widehat{\eta})$ and $\widehat{\beta}_{MLE}=n^{-1}\sum_{i=1}^{n}\widehat{F}_{T|W}(t|W_i)$. Then we solve (3) to obtain $\widehat{\beta}(D)$. Note that $\widehat{\beta}_{MLE}$ is the MLE of β under model $F_{T|W}(u|W;\eta)$. Thus, $\widehat{\beta}(D)$ is CAN, well behaved in finite samples and generally more efficient than $\widehat{\beta}(D)$ if the working model $F_{T|W}(u|W;\eta)$ holds. However $\widehat{\beta}_{MLE}$, in contrast to the doubly robust estimator $\widehat{\beta}(D)$, is inconsistent if the model $F_{T|W}(u|W;\eta)$ is mispecified.

2.3 A survey of applications of the AIPW methodology under CAR in survival analysis problems

The book of van der Laan and Robins, 2003, contains a comprehensive treatment of the AIPW methodology for estimation in censored and missing data models and in counterfactual models for causal inference under the CAR assumption. Here we review the literature on AIPW estimation restricting attention to censored failure time data in non-counterfactual models.

Robins and Rotnitzky, 1992, Robins, 1993 and Robins and Finkelstein, 2000, constructed locally efficient AIPW estimators of the survival distribution of T and of regression parameters of Cox proportional hazards models and accelerated failure time models for the conditional distribution of T given baseline covariates. Robins,1996, constructed AIPW estimators of median regression models for right censored failure time data and Robins, Rotnitzky and Zhao, 1994, and Nan, Emond and Wellner, 2004, described AIPW estimation of Cox proportional hazards regression parameters with missing covariates. Robins, Rotnitzky and Bonetti, 2001, described AIPW estimation of a failure time distribution under double-sampling with follow-up of dropouts. Hu and Tsiatis, 1996, used the AIPW methodology to construct estimators of a survival function from right censored data subject to reporting delays. Zhao and Tsiatis, 1997, 1999, 2000, and Van der Laan and Hubbard, 1999, constructed AIPW estimators of the quality of life adjusted survival time distribution from right censored data. Bang and Tsiatis, 2000, 2002, and Strawderman, 2000, derived respectively AIPW estimators of a median regression model for medical costs from right censored data and of the mean of an increasing stochastic process. Van der Laan, Hubbard and Robins, 2002, and Quale, van der Laan and Robins, 2003, constructed locally efficient AIPW estimators of a multivariate survival function when failure times are subject to a common censoring time and to a failure-time-specific censoring respectively. In the same setting, Keles, van der Laan and Robins, 2003, derived AIPW estimators that are easier to compute and almost as efficient than the Quale et al. estimators.

Robins and Rotnitzky, 1992, restricted their investigation to data configurations for which the full data X has a positive probability of being completely observed. Their work was later extended to censored data structures under the CAR assumption in which X is never completely observed. These extensions include the estimation of the marginal survival function of T and of regression parameters of an accelerated failure time model for the law of T given baseline covariates $V(0)$ from current status and/or interval censored data when the intensity function for monitoring whether T has occurred by t depends on the observed covariate history $\overline{V}(t)$ (van der Laan and Hubbard, 1997, van der Laan and Robins, 1998, 2002).

3 AIPW estimation without the CAR assumption

The CAR assumption (1) implies that the data $V(t), t \geq 0$ include all the time dependent and time independent prognostic factors for failure that also predict censoring. In most studies, however, data are typically available on some but not all joint prognostic factors for censoring and survival and hence CAR fails. Scharfstein, Robins, Eddings and Rotnitzky, 2001, and Scharfstein and Robins, 2002, have extended the AIPW methodology to allow estimation of the marginal survivor function at a fixed $t, \beta = \Pr(T > t)$, of a discrete and continuous failure time T respectively under non-CAR models. Their work was an extension to the analysis of failure time data of the AIPW methodology in non-CAR models for non-failure time endpoints derived in a series of papers by Rotnitzky and Robins,1997, Rotnitzky, Shcarfstein and Robins, 1998, Robins, Rotnitzky and Scharfstein, 2000, and Scharfstein, Rotnitzky and Robins, 1999. This methodology allows the analyst to appropriately adjust for informative censoring due to measured prognostic factors while simultaneously quantifying the sensitivity of inference to non-identifying assumptions concerning residual dependence between the failure time and censoring due to unmeasured factors. For continuous failure time data, their approach relies on the assumption that the censoring mechanism follows the model

$$\lambda_C\left(u|\overline{V}(u), T\right) = \lambda_{0,C}\left(u, \overline{V}(u)\right) \exp\left\{q\left(u, \overline{V}(u), T\right)\right\} \text{ for all } u \geq 0 \tag{4}$$

where $\lambda_{0,C}\left(u, \overline{V}(u)\right)$ is an unknown function and $q\left(u, \overline{V}(u), T\right)$ is a user-specified (i.e. known) function. The function $q\left(u, \overline{V}(u), T\right)$ quantifies, for those who remain at risk at time u, the dependence measured on the hazard ratio scale between T and censoring just after u after having adjusted for prognostic factors $\overline{V}(u)$. The choice $q\left(u, \overline{V}(u), T\right) = 0$ corresponds to an assumption slightly less stringent than the CAR assumption (1). Scharfstein and Robins, arguing like in Scharfstein, Rotnitzky and Robins, 1999, showed that their model, like models proposed in the competing risks literature without auxiliary data $V(t)$ (Fisher and Kanarek, 1974, Slud and Rubinstein, 1983, Klein and Moeschberger, 1988, Zheng and Klein, 1994, 1995), is a non-parametric, just identified model for the law of the observed data Y; that is, the function $q\left(u, \overline{V}(u), T\right)$ is not identified, but, once specified, the survivor parameter β is identified but the law of Y is not restricted. Following the lead in the competing risks literature, these authors recommended drawing inference about β by varying $q\left(u, \overline{V}(u), T\right)$ over a plausible range and described a useful parameterization of this function for conducting such analysis.

Model (4) is stringent enough to allow identification for β. However, as in the CAR model, the model is not stringent enough to allow well behaved estimation of β in finite samples when the covariate process is high dimensional because the function $\lambda_{0,C}\left(u,\overline{V}\left(u\right)\right)$ cannot estimated well due to the curse of dimensionality. In order to reduce the dimension of the unknown function $\lambda_{0,C}\left(u,\overline{V}\left(u\right)\right)$, Scharfstein and Robins, 2002, assumed a lower dimensional model for $\lambda_{0,C}\left(u,\overline{V}\left(u\right)\right)$ of the form

$$\lambda_{0,C}\left(u,\overline{V}\left(u\right)\right)=\lambda_{0,C}^{*}\left(u\right)\exp\left\{\gamma' w\left(u,\overline{V}\left(u\right)\right)\right\}$$

where $\lambda_{0,C}^{*}\left(u\right)$ and γ are unknown and $w\left(u,\overline{V}\left(u\right)\right)$ is a user specified function and described AIPW estimators of β under this model. Unlike the CAR model, this model does not admit doubly robust estimators (Robins and Rotnitzky, 2001).

References

[1] Bang H. and Tsiatis AA. (2000) Estimating medical cost with censored data. *Biometrika*, **87**:329-343.

[2] Bang H. and Tsiatis AA. (2002) Median regression with censored medical cost data, *Biometrics*, 58, 643-650.

[3] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993), *Efficient and Adaptive Inference in Semiparametric Models*, Baltimore, MD: Johns Hopkins University Press.

[4] Fisher, L. and Kanarek, P.(1974). Presenting censored survival data when censoring and survival times may not be independent. In Reliability and Biometry: Statistical Analysis of Lifelength. *SIAM*.

[5] Gill, R.D., van der Laan, M.J., and Robins, J.M. (1997). Coarsening at random: characterizations, conjectures and counterexamples. *Proceedings of the First Seattle Symposium on Survival Analysis*, pp. 255-294.

[6] Heitjan, D.F., and Rubin, D.B. (1991), Ignorability and Coarse Data, *The Annals of Statistics*, 19, 2244-2253.

[7] Hu, P.H. and Tsiatis, A.A. (1996). Estimating the survival function when ascertainment of vital status is subject to delay. *Biometrika*. 83: 371-380.

[8] Jacobsen, M. and Keiding, N. (1995). Coarsening at random in general samoke spaces and random censoring in continuous time. *Annals of Statistics,* 23, 774-786.

[9] Keles, S., van der Laan, M.J. and Robins, J. (2003). Estimation of the bivariate survival function in the presence of time dependent covariates. To appear in the *Survival Analysis Volume of the Handbook of Statistics*. C.R. Rao and Balakrishnan, N. editors.

[10] Keiding, N., Holst, C. and Green, A. (1989). Retrospective estimation of diabetes incidence from information in a current prevalent population and historical mortality. *American Journal of Epidemiology*, 130:588-600.

[11] Klein JP, Moeschberger ML. (1998) Bounds on net survival probabilities for dependent competing risks. *Biometrics*, **44**:529–538.

[12] Koul, H., Susarla, V., and van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Annals of Statistics*, 9:1276-88.

[13] Nan, B., Emond, M. and Wellner, J. (2004). Information bounds for Cox regression models with missing data. *The Annals of Statistitcs*. 32:2.

[14] Newey, W.K. (1990). Semiparametric Efficiency Bounds. *Journal of Applied Econometrics*, 5, 99-135.

[15] Quale, van der Laan and Robins (2003). Locally efficient estimation with bivariate right censored data. University of California, Berkeley. Department of Statistics. Technical Report.

[16] Robins, JM. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section*, American Statistical Association. Alexandria, VA. 24-33.

[17] Robins, JM. (1996). Locally efficient median regression with random censoring and surrogate markers. Proceedings of the 1994 Conference on Lifetime Data Models in Reliability

and Survival Analysis, Boston, MA. In: *Lifetime Data: Models in Reliability and Survival Analysis* NP Jewell et al., Eds., Kluwer, 1996, 263-274.

[18] Robins, J.M. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. *Proceedings 1999 Joint Statistical Meetings.*

[19] Robins, J.M., and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology - Methodological Issues.* Eds: Jewell N., Dietz K., Farewell V. Boston, MA: Birkhäuser. pp. 297-331.

[20] Robins, J., Rotnitzky, A. and Zhao, L.P. (1994). Estimation of regression coefficients when some of the regressors are nor always observed. *Journal of the American Statistical Association*, **89**:846–866.

[21] Robins, J.M. and Ritov, Y. (1997). A curse of dimensionality appropriate (CODA) asymptotic theory for semiparametric models. *Statistics in Medicine,* 16:285-319.

[22] Robins JM, Rotnitzky A, Scharfstein DO. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: *Statistical Models for Epidemiology, the Environment, and Clinical Trials.* E Halloran and D Berry, editors. pp 1-95.

[23] Robins, J.M., Rotnitzky, A., and van der Laan, M.J. (2000). Discussion of "On profile likelihood" by Murphy and van der Vaart. *Journal of the American Statistical Association*, 95:477-482.

[24] Robins, J.M. and Finkelstein, D. (2000). Correcting for non-compliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, 56(3):779-788.

[25] Robins J, Rotnitzky A, Bonetti M. (2001) Discussion of the paper by Frangakis C and Rubin D. "A note on addressing an idiosyncrasy in estimating survival curves using double-sampling in the presence of self-selected right censoring." *Biometrics*, **57** :343–347.

[26] Robins, J. and Rotnitzky, A. (2001). Discussion of the paper by Bickel and Kwon, Inference for semiparametric models: some questions and an answer. *Statistica Sinica,* **11(4):863-960.**

[27] Rotnitzky A. and Robins, J. (1997). Analysis of semi-parametric regression models with non-ignorable non-response. *Statistics in Medicine,* **16:81-102.**

[28] Rotnitzky, A., Robins, J.M. and Scharfstein, D. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statstical Association*, 93(444):1321-1339.

[29] Scharfstein DO, Rotnitzky A, Robins JM. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association*, **94(448)**:1096–1120.

[30] Scharfstein DO, Robins JM, Eddings W, Rotnitzky A. (2001). Inference in randomized studies with informative censoring and discrete time-to-event endpoints. *Biometrics*, **57(2)**:404–413.

[31] Scharfstein DO and Robins JM. (2002). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika.* 89: 617-634.

[32] Slud EV, Rubinstein LV. (1983). Dependent competing risks and summary survival curves. *Biometrika*, **70**:643–649.

[33] Strawderman R.L. (2000). Estimating the mean of an increasing stochastic process at a censored stopping time. *Journal of the American Statistical Association*, **95**:1192–1208.

[34] van der Laan MJ, Hubbard H. (1997) Estimation of interval censored data and covariates. *Lifetime Data Analysis*, **3**:77–91.

[35] van der Laan MJ, Robins J. (1998) Locally efficient estimation with current status data and time dependent covariates. *Journal of the American Statistical Association*, **93**:693–701.

[36] van der Laan MJ, Hubbard H. (1999). Locally efficient estimation of the quality adjusted lifetime distribution with right-censored data and covariates. *Biometrics,* **55**:530–536.

[37] van der Laan MJ, Hubbard H, Robins J. (2002) Locally efficient estimation of a multivariate survival function in longitudinal studies. *Journal of the American Statistical Association.* 97: 494-507.

[38] van der Laan MJ, Robins J. (2003) *Unified methods for censored and longitudinal data and causality.* Springer Verlag. New York.

[39] van der Vaart, A.W. (1988). *Statistical Estimation in Large Parameter Spaces.* CWI Tract, Centre for Mathematics and Computer Science, Amsterdam.

[40] van der Vaart, A.W. (1991). On differentiable functionals. *Annals of Statistics.* 19, 178-204.

[41] Zhao H, and Tsiatis AA. (1997). A Consistent Estimator for the Distribution of Quality Adjusted Survival Time. *Biometrika*, 1997; **84**:339–348.

[42] Zhao H, and Tsiatis AA. (1999) Efficient Estimation of the Distribution of Quality Adjusted Survival Time. *Biometrics*, **55**:1101–1107.

[43] Zhao H, and Tsiatis AA. (2000) Estimating Mean Quality Adjusted Lifetime With Censored Data. *Sankhya*, **62, Series B(1)** :175–188.

[44] Zheng M, Klein JP. (1994) A self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula. *Communications in Statistics, Part A - Theory and Methods*, 1994; **23**:2299–2311.

[45] Zheng M, Klein JP. (1995) Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, **82**:127–138.