Nonparametric inference for assessing treatment efficacy in randomized clinical trials with a time-to-event outcome and all-or-none compliance

BY ROBERT M. ELASHOFF, GANG LI AND YING ZHOU

*Department of Biostatistics, School of Public Health, University of California Los Angeles,
Los Angeles, California 90095, U.S.A.*

relashof@biomath.ucla.edu vli@ucla.edu zhying1000@gmail.com

SUMMARY

To evaluate the biological efficacy of a treatment in a randomized clinical trial, one needs to compare patients in the treatment arm who actually received treatment with the subgroup of patients in the control arm who would have received treatment had they been randomized into the treatment arm. In practice, subgroup membership in the control arm is usually unobservable. This paper develops a nonparametric inference procedure to compare subgroup probabilities with right-censored time-to-event data and unobservable subgroup membership in the control arm. We also present a procedure to estimate the onset and duration of treatment effect. The performance of our method is evaluated by simulation. An illustration is given using a randomized clinical trial for melanoma.

Some key words: Biological efficacy; Censoring; Counting process; Martingale; Noncompliance; Survival probability.

1. INTRODUCTION

In randomized clinical trials, subjects often fail to comply with their assigned treatment regimen. The intent-to-treat analysis is a standard method for primary analysis of randomized trials that compares the treatment and control based on the initial randomization regardless of whether the subjects actually received their assigned treatment or not. However, in the presence of non-compliance, the intent-to-treat analysis estimates the programmatic effectiveness, not the biological efficacy of the treatment (Sommer & Zeger, 1991; Hollis & Campbell, 1999; Lachin, 2000).

Consider a randomized clinical trial with all-or-none compliance in the treatment arm, but no noncompliance in the control arm, where all-or-none compliance means that subjects either fully comply or fully do not comply with their assigned treatment regimen. To evaluate the treatment efficacy, one needs to compare the compliers in the treatment arm with the latent subgroup of patients in the control arm who would have received the treatment if they were offered the treatment. However, the membership of the latent compliance subgroup in the control arm is unobservable. For example, in the Eastern Cooperative Oncology Group trial E9288 (Kemeny et al., 2002), patients with liver metastases of colorectal cancer were randomized to receive surgical resection alone or surgical resection followed by chemotherapy. All patients received surgical resection, but only a portion of those randomized to the chemotherapy arm actually received the postoperative chemotherapy for reasons related to survival. To assess the biological efficacy of chemotherapy, one should compare patients in the chemotherapy arm who have actually received chemotherapy with those in the no-chemotherapy arm who would have received

chemotherapy had they been assigned to receive postoperative chemotherapy. However, the latent chemotherapy-compliance status is not observable in the no-chemotherapy arm. Another example is Zelen's (1979, 1990) single-consent design, in which subjects randomized to the control arm must receive the standard care, and those to the treatment arm can choose either the experimental treatment or the standard care.

This problem also arises from subgroup analysis when the subgroup status is identified by a diagnostic test that is performed in the treatment arm but not in the control arm. One example is the Multicenter Selective Lymphadenectomy Trial I (Morton et al., 2006). In this trial, newly diagnosed melanoma patients were randomized to a sentinel-node biopsy arm or a nodal observation arm. All patients underwent wide excision of the primary melanoma. In the biopsy arm, sentinel-node biopsy was performed on patients to identify the presence of sentinel-node metastases. An immediate complete lymphadenectomy was then performed for patients whose sentinel-node biopsies were positive for metastases. For all other patients in the trial, delayed complete lymphadenectomy was performed only when nodal recurrences became clinically detectable. An interesting question is whether the trial provides sufficient evidence that immediate complete lymphadenectomy improves survival relative to delayed complete lymphadenectomy in patients with sentinel-node metastases. The design of this trial does not facilitate a direct comparison since the sentinel-node status is unknown in the nodal observation arm.

There is a large literature on treatment noncompliance. For instance, Sommer & Zeger (1991) studied the problem with a dichotomous outcome and all-or-none compliance. Their approach has been extended to handle ordered categorical compliance (Goetghebeur & Molenberghs, 1996), all-or-none compliance with contamination (Cuzick et al., 1997) and clustered binary outcomes (Albert, 2002). Several methods are available for a time-to-event outcome with right-censored data. Robins & Tsiatis (1991) proposed a structural failure time model based on potential outcomes. Loeys & Goetghebeur (2003), Loeys et al. (2005) and Cuzick et al. (2007) proposed methods based on proportional hazards models. Frangakis & Rubin (1999) studied a nonparametric procedure to evaluate treatment efficacy in a randomized trial with missing outcomes and all-or-none noncompliance and discussed an extension to right-censored time-to-event data. Loeys & Goetghebeur (2003) proposed a nonparametric plug-in estimator of the survival function for compliers in the control arm, but its large-sample properties have not been studied.

The purpose of this paper is to develop a nonparametric inference method to assess treatment efficacy in randomized clinical trials with a right-censored time-to-event outcome and all-or-none compliance. We derive large sample properties of the estimated subgroup survival probabilities, which lead to an asymptotic inference procedure to compare a survival probability for compliers in the treatment arm with that for the latent compliers in the control arm. We also provide a procedure to determine the onset and duration of the treatment effect.

2. METHODS

2.1. Notation and assumptions

Suppose that n_2 subjects are randomized to the treatment arm and n_1 subjects to the control arm. Let $n = n_1 + n_2$. In the treatment arm, one observes n_2 independent and identically distributed triplets $(X_{2i}, \delta_{2i}, g_{2i})$ ($i = 1, \dots, n_2$), where for the i th subject the observation time $X_{2i} = \min(T_{2i}, C_{2i})$ is the minimum of a nonnegative continuous survival time T_{2i} and a censoring time C_{2i} , $\delta_{2i} = I(T_{2i} \leq C_{2i})$ is a failure indicator, and g_{2i} is a binary compliance indicator, 1 for compliance. In the control arm, one observes n_1 independent and identically distributed pairs (X_{1i}, δ_{1i}) ($i = 1, \dots, n_1$), where for subject i , $X_{1i} = \min(T_{1i}, C_{1i})$, $\delta_{1i} = I(T_{1i} \leq C_{1i})$ and

Table 1. Notation of survival functions by study arms and compliance status

	Treatment	Control
Compliers ($g_i = 1$)	$S_2(t) = \text{pr}(T_{2i} > t \mid g_{2i} = 1)$	$S_1(t) = \text{pr}(T_{1i} > t \mid g_{1i} = 1)$
Noncompliers ($g_i = 0$)	$G_2(t) = \text{pr}(T_{2i} > t \mid g_{2i} = 0)$	$G_1(t) = \text{pr}(T_{1i} > t \mid g_{1i} = 0)$
Overall	—	$H_1(t) = \text{pr}(T_{1i} > t)$

the binary compliance indicator g_{1i} is not observed. Table 1 defines the subgroup survival functions by study arms and compliance status.

Let $\Lambda_1(t)$, $\Lambda_2^S(t)$ and $\Lambda_2^G(t)$ denote the cumulative hazard functions of $H_1(t)$, $S_2(t)$ and $G_2(t)$, respectively. In addition, let $D_1(t) = \text{pr}(C_{1i} > t)$, $D_2^S(t) = \text{pr}(C_{2i} > t \mid g_{2i} = 1)$ and $D_2^G(t) = \text{pr}(C_{2i} > t \mid g_{2i} = 0)$.

The following assumptions are made throughout the paper.

Assumption 1. The compliance proportion is independent of randomization, i.e., $\text{pr}(g_{1i} = 1) = \text{pr}(g_{2i} = 1) \equiv p$.

Assumption 2. The survival function of noncompliers is independent of randomization, i.e., $G_1(t) = G_2(t)$, for all t .

Assumption 3. In the control arm, the censoring time is independent of the survival time. In the treatment arm, the censoring time is independent of the survival time conditional on the compliance status.

2.2. The estimator

The overall survival function for the control arm is $H_1(t) = \text{pr}(g_{1i} = 1)S_1(t) + \text{pr}(g_{1i} = 0)G_1(t)$. This, together with Assumptions 1 and 2, implies that

$$S_1(t) = \{H_1(t) - (1 - p)G_2(t)\}/p,$$

which leads to the plug-in estimator (Loeys & Goetghebeur, 2003)

$$\hat{S}_1(t) = \{\hat{H}_1(t) - (1 - \hat{p})\hat{G}_2(t)\}/\hat{p}. \quad (1)$$

Here $\hat{H}_1(t) = \prod_{X_{1i:n_1} \leq t} \{1 - 1/(n_1 - i + 1)\}^{\delta_{1i:n_1}}$ is the Kaplan & Meier (1958) estimator of $H_1(t)$, where $0 \leq X_{11:n_1} \leq \dots \leq X_{1n_1:n_1}$ are the ordered observation times in the control arm and $\delta_{1i:n_1}$ is the failure indicator for $X_{1i:n_1}$, $\hat{G}_2(t)$ is the Kaplan–Meier estimator of $G_2(t)$ based on the noncompliers in the treatment arm and $\hat{p} = \sum_{i=1}^{n_2} g_{2i}/n_2$.

Let $\hat{S}_2(t)$ be the Kaplan–Meier estimator of $S_2(t)$ based on the compliers ($g_{2i} = 1$) in the treatment arm. We estimate $S_2(t) - S_1(t)$, a measure of treatment efficacy, by $\hat{S}_2(t) - \hat{S}_1(t)$.

Remark 1. Frangakis & Rubin (1999, § 5) derived a different estimator of $S_1(t)$ that is consistent for $S_1(t)$ under independent censoring conditional on compliance status in both arms. Because compliance status is unobserved in the control arm, the conditional independent censoring assumption in the control arm is untestable. Furthermore, the plug-in estimator defined in (1) may not be consistent under the assumptions of Frangakis & Rubin (1999) since the conditional independent censoring assumption in the control arm renders the Kaplan–Meier estimator $\hat{H}_1(t)$ generally inconsistent for $H_1(t)$. We make different assumptions in Assumption 3 that the censoring time is independent of the survival time conditional on the compliance status in the

treatment arm, but unconditionally independent in the control arm. Thus, under Assumption 3, $\hat{H}_1(t)$ is consistent for $H_1(t)$ and $\hat{G}_2(t)$ is consistent for $G_2(t)$. Consequently, the plug-in estimator $\hat{S}_1(t)$ is consistent for $S_1(t)$.

2.3. Asymptotic properties

The counting process and martingale theory are used to study the asymptotic properties of $\hat{S}_1(t)$ and $\hat{S}_2(t) - \hat{S}_1(t)$. For the control arm, define the at-risk processes

$$Y_{1i}(t) = I(X_{1i} \geq t), \quad Y_1(t) = \sum_{i=1}^{n_1} Y_{1i}(t),$$

and counting processes

$$N_{1i}(t) = I(X_{1i} \leq t, \delta_{1i} = 1), \quad N_1(t) = \sum_{i=1}^{n_1} N_{1i}(t).$$

Then $M_{1i}(t) = N_{1i}(t) - \int_0^t Y_{1i}(u) d\Lambda_1(u)$ ($i = 1, \dots, n_1$), are orthogonal locally integrable martingales with respect to the filtration $\mathcal{F}_1(t) = \sigma[\{N_{1i}(u), Y_{1i}(u)\}, i = 1, \dots, n_1 : 0 \leq u \leq t]$, the failure and censoring histories up to time t .

Similarly, let $Y_{2i}(t)$ and $N_{2i}(t)$ denote the corresponding at-risk process and counting process for subject i in the treatment arm. For the compliance subgroup in the treatment arm, define $Y_2^S(t) = \sum_{i=1}^{n_2} g_{2i} Y_{2i}(t)$, $N_2^S(t) = \sum_{i=1}^{n_2} g_{2i} N_{2i}(t)$ and $M_2^S(t) = \sum_{i=1}^{n_2} g_{2i} M_{2i}^S(t)$, where $M_{2i}^S(t) = N_{2i}(t) - \int_0^t Y_{2i}(u) d\Lambda_2^S(u)$. For the noncompliance subgroup in the treatment arm, define $Y_2^G(t) = \sum_{i=1}^{n_2} (1 - g_{2i}) Y_{2i}(t)$, $N_2^G(t) = \sum_{i=1}^{n_2} (1 - g_{2i}) N_{2i}(t)$ and $M_2^G(t) = \sum_{i=1}^{n_2} (1 - g_{2i}) M_{2i}^G(t)$, where $M_{2i}^G(t) = N_{2i}(t) - \int_0^t Y_{2i}(u) d\Lambda_2^G(u)$.

Let

$$\hat{\Lambda}_1(t) = \int_0^t \frac{dN_1(u)}{Y_1(u)}, \quad \hat{\Lambda}_2^S(t) = \int_0^t \frac{dN_2^S(u)}{Y_2^S(u)}, \quad \hat{\Lambda}_2^G(t) = \int_0^t \frac{dN_2^G(u)}{Y_2^G(u)}$$

be the Nelson–Aalen estimators of $\Lambda_1(t)$, $\Lambda_2^S(t)$ and $\Lambda_2^G(t)$.

The following lemma is needed to derive the joint distribution of $\hat{H}_1(t)$, $\hat{S}_2(t)$, $\hat{G}_2(t)$ and $\hat{p}$.

LEMMA 1. Assume that $\Lambda_1(\tau) < \infty$, $\Lambda_2^S(\tau) < \infty$, $\Lambda_2^G(\tau) < \infty$, $D_1(\tau) > 0$, $D_2^S(\tau) > 0$ and $D_2^G(\tau) > 0$, where τ is usually the time when data collection ends. Assume $n_1/n \rightarrow \rho_1$ and $n_2/n \rightarrow \rho_2$ for $0 < \rho_1, \rho_2 < 1$ as $n \rightarrow \infty$. Then, for any t in $(0, \tau]$,

$$n^{1/2} \begin{pmatrix} \hat{H}_1(t) - H_1(t) \\ \hat{S}_2(t) - S_2(t) \\ \hat{G}_2(t) - G_2(t) \\ \hat{p} - p \end{pmatrix} \rightarrow N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma(t) \right)$$

in distribution, as $n \rightarrow \infty$, where the variance–covariance matrix is

$$\Sigma(t) = \begin{pmatrix} H_1^2(t) \rho_1^{-1} \sigma_{11}(t) & 0 & 0 & 0 \\ 0 & S_2^2(t) \rho_2^{-1} \sigma_{22}(t) & 0 & -S_2(t) \rho_2^{-1} \sigma_{24}(t) \\ 0 & 0 & G_2^2(t) \rho_2^{-1} \sigma_{33}(t) & -G_2(t) \rho_2^{-1} \sigma_{34}(t) \\ 0 & -S_2(t) \rho_2^{-1} \sigma_{24}(t) & -G_2(t) \rho_2^{-1} \sigma_{34}(t) & \rho_2^{-1} p(1 - p) \end{pmatrix},$$

with

$$\begin{aligned}\sigma_{11}(t) &= E \left[\left\{ \int_0^t \frac{dM_{1i}(u)}{H_1(u-)D_1(u-)} \right\}^2 \right], \quad \sigma_{22}(t) = E \left[\left\{ \int_0^t \frac{dM_{2i}^S(u)g_{2i}}{pS_2(u-)D_2^S(u-)} \right\}^2 \right], \\ \sigma_{33}(t) &= E \left[\left\{ \int_0^t \frac{dM_{2i}^G(u)(1-g_{2i})}{(1-p)G_2(u-)D_2^G(u-)} \right\}^2 \right], \\ \sigma_{24}(t) &= E \left\{ (g_{2i} - p) \int_0^t \frac{dM_{2i}^S(u)g_{2i}}{pS_2(u-)D_2^S(u-)} \right\}, \\ \sigma_{34}(t) &= E \left\{ (g_{2i} - p) \int_0^t \frac{dM_{2i}^G(u)(1-g_{2i})}{(1-p)G_2(u-)D_2^G(u-)} \right\}.\end{aligned}$$

Furthermore, $\sigma_{11}(t)$, $\sigma_{22}(t)$, $\sigma_{33}(t)$, $\sigma_{24}(t)$ and $\sigma_{34}(t)$ can be consistently estimated by

$$\begin{aligned}\hat{\sigma}_{11}(t) &= n_1 \sum_{i=1}^{n_1} \left\{ \int_0^t \frac{d\hat{M}_{1i}(u)}{Y_1(u)} \right\}^2, \quad \hat{\sigma}_{22}(t) = n_2 \sum_{i=1}^{n_2} \left\{ \int_0^t \frac{d\hat{M}_{2i}^S(u)g_{2i}}{Y_2^S(u)} \right\}^2, \\ \hat{\sigma}_{33}(t) &= n_2 \sum_{i=1}^{n_2} \left\{ \int_0^t \frac{d\hat{M}_{2i}^G(u)(1-g_{2i})}{Y_2^G(u)} \right\}^2, \quad \hat{\sigma}_{24}(t) = \sum_{i=1}^{n_2} (g_{2i} - \hat{p}) \int_0^t \frac{d\hat{M}_{2i}^S(u)g_{2i}}{Y_2^S(u)}, \\ \hat{\sigma}_{34}(t) &= \sum_{i=1}^{n_2} (g_{2i} - \hat{p}) \int_0^t \frac{d\hat{M}_{2i}^G(u)(1-g_{2i})}{Y_2^G(u)},\end{aligned}$$

where $\hat{M}_{1i}(t) = N_{1i}(t) - \int_0^t Y_{1i}(u) d\hat{\Lambda}_1(u)$, $\hat{M}_{2i}^S(t) = N_{2i}(t) - \int_0^t Y_{2i}(u) d\hat{\Lambda}_2^S(u)$ and $\hat{M}_{2i}^G(t) = N_{2i}(t) - \int_0^t Y_{2i}(u) d\hat{\Lambda}_2^G(u)$.

The joint limiting distribution of $\hat{S}_1(t)$ and $\hat{S}_2(t)$ is stated below.

THEOREM 1. Assume that the assumptions of Lemma 1 hold. At a given time-point t in $(0, \tau]$,

$$n^{1/2} \begin{pmatrix} \hat{S}_1(t) - S_1(t) \\ \hat{S}_2(t) - S_2(t) \end{pmatrix} \rightarrow N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} v_{11}(t) & v_{12}(t) \\ v_{12}(t) & v_{22}(t) \end{pmatrix} \right\}$$

in distribution, as $n \rightarrow \infty$, $n_1/n \rightarrow \rho_1$ and $n_2/n \rightarrow \rho_2$, where

$$\begin{aligned}v_{11}(t) &= \frac{H_1^2(t)}{p^2} \rho_1^{-1} \sigma_{11}(t) + \frac{(1-p)^2 G_2^2(t)}{p^2} \rho_2^{-1} \sigma_{33}(t) + \frac{\{H_1(t) - G_2(t)\}^2}{p^4} \rho_2^{-1} \sigma_{44} \\ &\quad - 2 \frac{(1-p)G_2(t)\{H_1(t) - G_2(t)\}}{p^3} \rho_2^{-1} \sigma_{34}(t),\end{aligned}\tag{2}$$

$$v_{22}(t) = S_2^2(t) \rho_2^{-1} \sigma_{22}(t),\tag{3}$$

$$v_{12}(t) = \frac{S_2(t)\{H_1(t) - G_2(t)\}}{p^2} \rho_2^{-1} \sigma_{24}(t).\tag{4}$$

The proofs of Lemma 1 and Theorem 1 are provided in the Appendix.

2.4. Pointwise confidence intervals for survival probabilities

Let $\hat{v}_{11}(t)$, $\hat{v}_{22}(t)$ and $\hat{v}_{12}(t)$ be the consistent estimates of $v_{11}(t)$, $v_{22}(t)$ and $v_{12}(t)$ obtained by replacing the theoretical quantities in (2)–(4) with their consistent estimators. It follows from Theorem 1 that the $100(1 - \alpha)\%$ confidence intervals for $S_1(t)$ and $S_2(t)$ at a given $t \in [0, \tau]$ are given by

$$\hat{S}_1(t) \pm z_{1-\alpha/2} \left\{ \frac{\hat{v}_{11}(t)}{n} \right\}^{1/2}, \quad \hat{S}_2(t) \pm z_{1-\alpha/2} \left\{ \frac{\hat{v}_{22}(t)}{n} \right\}^{1/2},$$

respectively, where $z_{1-\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

Furthermore, a $100(1 - \alpha)\%$ confidence interval for $S_2(t) - S_1(t)$ is given by

$$\{\hat{S}_2(t) - \hat{S}_1(t)\} \pm z_{1-\alpha/2} n^{-1/2} \hat{\sigma}(t),$$

where $\hat{\sigma}^2(t) = \hat{v}_{11}(t) + \hat{v}_{22}(t) - 2\hat{v}_{12}(t)$.

2.5. Onset and duration of treatment effect

In practice, one is often interested in the estimation of the time interval in which $S_2(t)$ exceeds $S_1(t)$. Below we provide a procedure to determine the onset and duration of the time interval with a confidence level $1 - \alpha$. The procedure is based on ideas from Berger & Boos (1999).

Step 1. Construct a one-sided, $\alpha/2$ -level test of $H_{0t} : S_2(t) - S_1(t) = 0$ versus $H_{at} : S_2(t) - S_1(t) > 0$ for a given $t \in [0, \tau]$. Define $Z_t = \{\hat{S}_2(t) - \hat{S}_1(t)\} / \{n^{-1/2} \hat{\sigma}(t)\}$. We reject the null hypothesis H_{0t} if $Z_t > z_{1-\alpha/2}$.

Step 2. Choose starting value $t_s \in [0, \tau]$. If Z_{t_s} accepts H_{0t_s} , no confidence statement is made. If Z_{t_s} rejects H_{0t_s} , we test sequentially downward and upward from t_s . Let L be the last $t \leq t_s$ for which H_{0t} is rejected, and U be the last $t \geq t_s$ for which H_{0t} is rejected. Then $[L, U]$ is the largest interval containing t_s for which Z_t rejects H_{0t} for all $t \in [L, U]$.

THEOREM 2. Let $[L, U]$ be defined by the above algorithm. Then,

$$\text{pr}\{S_2(t) - S_1(t) > 0, t \in [L, U]\} \geq 1 - \alpha,$$

as $n \rightarrow \infty$, $n_1/n \rightarrow \rho_1$ and $n_2/n \rightarrow \rho_2$ for some constants $0 < \rho_1, \rho_2 < 1$.

As remarked by Berger & Boos (1999), the starting value t_s should be chosen before the study. Any starting point satisfying $L \leq t_s \leq U$ will lead to the same interval $[L, U]$. The choice of t_s requires some prior information about the time frame during which $S_2(t)$ is likely to be higher than $S_1(t)$. If prior knowledge is unavailable, one can repeat the procedure at k different starting points and use level $\alpha/(2k)$ for each one-sided test to achieve the overall confidence level $1 - \alpha$.

3. SIMULATION STUDIES

We carried out some simulations to investigate the finite sample performance of our method for different combinations of sample size, censoring rate and compliance proportion.

In the first simulation, the survival times were generated from Weibull distributions $S_1(t) = \exp(-t^{0.8})$, $S_2(t) = \exp(-t^{0.75}/2)$ and $G_1(t) = G_2(t) = \exp(-t^{0.7}/4)$. The censoring times were generated from an exponential distribution with hazard rate λ chosen to give a prespecified overall censoring rate. For each simulation, 1000 Monte Carlo samples were generated.

Table 2. Performance of $\hat{S}_1(t)$, estimated standard error and confidence interval

p	CR	$S_1(t)$	$n_1 = n_2 = 100$				$n_1 = n_2 = 500$			
			$\hat{S}_1(t)$	SE	$\hat{SE}$	ACP	$\hat{S}_1(t)$	SE	$\hat{SE}$	ACP
0.7	0.3	0.50	0.50	0.08	0.08	94.2	0.50	0.04	0.04	94.8
		0.25	0.25	0.09	0.09	94.7	0.25	0.04	0.04	94.9
	0.5	0.50	0.50	0.08	0.08	94.8	0.50	0.04	0.04	95.2
		0.25	0.25	0.10	0.10	94.6	0.25	0.04	0.04	95.6
	0.7	0.50	0.50	0.10	0.09	94.5	0.50	0.04	0.04	94.9
		0.25	0.24	0.15	0.13	90.9	0.25	0.06	0.06	95.1
0.5	0.3	0.50	0.50	0.12	0.11	94.5	0.50	0.05	0.05	96.0
		0.25	0.24	0.13	0.13	94.6	0.25	0.06	0.06	95.4
	0.5	0.50	0.50	0.12	0.12	94.1	0.50	0.05	0.05	95.9
		0.25	0.24	0.15	0.14	94.6	0.25	0.06	0.06	95.3
	0.7	0.50	0.50	0.14	0.13	94.3	0.50	0.06	0.06	95.4
		0.25	0.24	0.19	0.18	94.6	0.25	0.08	0.08	94.9
0.3	0.3	0.50	0.49	0.20	0.19	95.7	0.50	0.08	0.08	95.9
		0.25	0.24	0.23	0.23	96.1	0.25	0.10	0.10	95.8
	0.5	0.50	0.49	0.21	0.20	95.6	0.50	0.08	0.09	95.4
		0.25	0.23	0.25	0.24	95.3	0.25	0.10	0.11	95.4
	0.7	0.50	0.49	0.22	0.21	95.9	0.50	0.09	0.09	96.0
		0.25	0.23	0.30	0.29	95.7	0.25	0.13	0.13	96.2

$\hat{S}_1(t)$, the empirical mean of estimated survival probability; SE, the empirical standard error of $\hat{S}_1(t)$; $\hat{SE}$, the empirical mean of estimated standard error of $\hat{S}_1(t)$; ACP, the achieved coverage probability of the 95% confidence interval for $S_1(t)$; p , the compliance proportion; CR, the censoring rate.

Table 2 reports the empirical mean of the plug-in estimate $\hat{S}_1(t)$, the empirical standard error of $\hat{S}_1(t)$, the empirical mean of estimated standard error of $\hat{S}_1(t)$ and the achieved coverage probability of the 95% confidence interval for $S_1(t)$. For the moderate sample size $n_1 = n_2 = 100$, the plug-in estimator $\hat{S}_1(t)$ and its standard error have very small biases and the confidence interval has very small coverage probability error in almost all cases. The only exception is that when the censoring rate is 70%, the coverage rate, 90.9%, is notably lower than the nominal level 95% in the right tail, $S(t) = 0.25$. Hence, one should interpret the analysis results for the right tail with caution in the presence of heavy censoring. Furthermore, as the sample sizes increase to $n_1 = n_2 = 500$, the performance of our method is satisfactory in all cases considered.

A similar simulation was conducted under conditions that emulate the Multicenter Selective Lymphadenectomy Trial I, in which the overall censoring rate is 87%, the observed proportion of node-positive patients is 15% and the sample sizes are $n_1 = 500$ and $n_2 = 764$. In this simulation, the survival times were generated from Weibull distributions $S_1(t) = \exp(-t^{0.87}/48.3)$, $S_2(t) = \exp(-t^{0.87}/89.8)$ and $G_1(t) = G_2(t) = \exp(-t^{0.79}/117.2)$. The results are summarized in Table 3. It is seen that similar to the previous simulation, the plug-in estimate $\hat{S}_1(t)$ and its standard error estimate show small bias and the confidence interval has small coverage probability error except in the right tail.

We conducted more simulations under different scenarios that include different survival distributions and a null case where $S_1 = S_2$. The performance of $\hat{S}_2(t) - \hat{S}_1(t)$ was also investigated. The results were similar to those for $\hat{S}_1(t)$ and thus omitted. We also carried out a small sensitivity analysis of Assumption 2 and did not observe serious estimation bias and coverage error under moderate censoring when Assumption 2 is slightly violated.

Table 3. *Simulation designed to emulate the Multicenter Selective Lymphadenectomy Trial I*

n_1	n_2	p	CR	$S_1(t)$	$\hat{S}_1(t)$	SE	$\hat{SE}$	ACP
500	764	0.15	0.87	0.50	0.51	0.25	0.25	95.30
				0.25	0.22	0.63	0.55	91.70

p , the sample size; CR, the censoring rate; $\hat{S}_1(t)$, the empirical mean of estimated survival probability; SE, the empirical standard error of $\hat{S}_1(t)$; $\hat{SE}$, the empirical mean of estimated standard error of $\hat{S}_1(t)$; ACP, the achieved coverage probability of the 95% confidence interval for $S_1(t)$.

Table 4. *Melanoma-specific survival probabilities, and standard errors in parentheses, for patients with sentinel-node metastases*

	t (years)			
	1	2	2.5	3
Delayed complete lymphadenectomy $\hat{S}_1(t)$	0.90 (0.04)	0.75 (0.08)	0.65 (0.09)	0.64 (0.10)
Immediate complete lymphadenectomy $\hat{S}_2(t)$	0.97 (0.02)	0.88 (0.03)	0.86 (0.03)	0.83 (0.04)
Difference $\hat{S}_2(t) - \hat{S}_1(t)$	0.06 (0.04)	0.13 (0.08)	0.22 (0.10)	0.20 (0.11)

4. AN EXAMPLE

We applied our method to analyse the Multicenter Selective Lymphadenectomy Trial I described in § 1. Between January 1994 and March 2002, 769 patients were randomized to the sentinel-node biopsy arm and 500 patients to the nodal observation arm. Our analysis excludes five patients in the biopsy arm whose sentinel-node status was unavailable. Immediate complete lymphadenectomy was performed on 122 of 764 patients in the biopsy arm whose biopsies were positive for sentinel-node metastases. For other patients, delayed complete lymphadenectomy was performed upon clinically observable nodal relapse.

To apply our method, it is important that Assumption 2 in § 2.1 is reasonable for this trial. Because removing the sentinel node among the node-negative patients is not expected to change their survival experience, it is reasonable to assume that $G_1(t) = G_2(t)$, where $G_1(t)$ and $G_2(t)$ represent the survival probabilities of node-negative patients in the observation arm and the biopsy arm, respectively.

We applied our method to estimate the benefit of immediate versus delayed complete lymphadenectomy on patients with sentinel-node metastases with respect to melanoma-specific survival, i.e., survival until death due to melanoma. Let $S_1(t)$ and $S_2(t)$ be the survival probabilities for patients with sentinel-node metastases in the observation arm and the biopsy arm, respectively. Table 4 reports the estimated survival probabilities $\hat{S}_1(t)$ and $\hat{S}_2(t)$, the estimated survival difference $\hat{S}_2(t) - \hat{S}_1(t)$ and their estimated standard errors at some given time-points. It is seen from Table 4 that the 95% confidence interval for $S_2(t) - S_1(t)$ is (0.02, 0.42) at 2.5 years, which implies that the 2.5-year melanoma-specific survival probability for patients with sentinel-node metastases in the biopsy arm is significantly higher than that of the observation arm at 0.05 significance level. Furthermore, using the procedure in § 2.5, we found that immediate complete lymphadenectomy significantly improves melanoma-specific survival relative to delayed complete lymphadenectomy for patients with sentinel-node metastases over the time interval [2.05, 2.97] years at 95% confidence level.

5. DISCUSSION

Most existing methods for assessing treatment efficacy compare either mean survival times (Robins & Tsiatis, 1991) or hazard rates (Loeys & Goetghebeur, 2003; Loeys et al., 2005; Cuzick et al., 2007) between study arms using parametric or semiparametric models. The difference in subgroup survival probabilities provides a useful alternative measure of treatment efficacy. Our method is fully nonparametric. If the proportional hazards assumption between $S_1(t)$ and $S_2(t)$ holds, our nonparametric method may not be as efficient as a proportional hazards model-based method. On the other hand, our method is more robust when the hazards of $S_1(t)$ and $S_2(t)$ are not proportional.

As illustrated in our simulations, caution is needed when drawing inference in the right tail of a survival distribution for moderate sample size, especially under heavy censoring and extremely unbalanced compliance proportions.

The focus of this paper is on comparison of subgroup survival probabilities at a fixed time-point based on the plug-in estimator $\hat{S}_1(t)$. As pointed out by Loeys & Goetghebeur (2003), $\hat{S}_1(t)$ is not a proper estimate of the entire survival curve because it is not monotonically non-increasing. One possible solution is to apply isotonic regression to the plug-in estimator to obtain a proper survival function. Properties of the resulting estimator, however, are difficult to study. An alternative approach is to consider nonparametric maximum likelihood estimation of S_1 .

ACKNOWLEDGEMENT

The authors are grateful to the editor, an associate editor and two referees for their insightful and constructive comments. We also thank Dr Donald L. Morton for providing the melanoma data used in the example. Gang Li's research was supported by a National Institutes of Health grant.

APPENDIX

Proofs

To prove the main results, we need the following lemma.

LEMMA A1. *Under the assumptions of Lemma 1, we have $\sup_{t \in [0, \tau]} |Y_2^S(t)/n_2 - y_2^S(t)| \rightarrow 0$, $\sup_{t \in [0, \tau]} |Y_2^G(t)/n_2 - y_2^G(t)| \rightarrow 0$ and $\sup_{t \in [0, \tau]} |Y_1(t)/n_1 - y_1(t)| \rightarrow 0$, in probability, where $y_2^S(t) = pS_2(t-)D_2^S(t-)$, $y_2^G(t) = (1-p)G_2(t-)D_2^G(t-)$ and $y_1(t) = H_1(t-)D_1(t-)$.*

Proof. For each $i = 1, \dots, n_2$, $Y_{2i} = I(X_{2i} \geq t)$ is a left-continuous nonincreasing random function. We assume that $S_2(t)$ and $D_2(t)$ are continuous functions. At a given time-point $t \in [0, \tau]$,

$$E\{g_{2i} Y_{2i}(t)\} = \text{pr}(g_{2i} = 1) \text{pr}\{Y_{2i}(t) = 1 | g_{2i} = 1\} = pS_2(t-)D_2^S(t-).$$

By the weak law of large numbers, at each given time-point t ,

$$\left| \frac{\sum_{i=1}^{n_2} g_{2i} Y_{2i}(t)}{n_2} - pS_2(t-)D_2^S(t-) \right| \rightarrow 0$$

in probability, as $n_2 \rightarrow \infty$. By the dominated convergence theorem,

$$E\{g_{2i} Y_{2i}(t+)\} = E \left\{ \lim_{k \rightarrow \infty} g_{2i} Y_{2i} \left(t + \frac{1}{k} \right) \right\} = \lim_{k \rightarrow \infty} E \left\{ g_{2i} Y_{2i} \left(t + \frac{1}{k} \right) \right\} = pS_2(t)D_2^S(t).$$

Thus, as $n_2 \rightarrow \infty$, $|\{\sum_{i=1}^{n_2} g_{2i} Y_{2i}(t+)/n_2 - p S_2(t) D_2^S(t)\}| \rightarrow 0$ in probability. Using arguments similar to those in the proof of [Chung \(2001, Theorem 5.5.1\)](#), it can be shown that $\sup_{t \in [0, \tau]} |Y_2^S(t)/n_2 - y_2^S(t)| \rightarrow 0$ in probability, as $n_2 \rightarrow \infty$, where $y_2^S(t) = p S_2(t-) D_2^S(t-)$.

The other results can be proved along the same lines. $\square$

Proof of Lemma 1. It is well known that $\hat{\Lambda}_1(t)$, $\hat{\Lambda}_2^S(t)$ and $\hat{\Lambda}_2^G(t)$ are consistent estimators of $\Lambda_1(t)$, $\Lambda_2^S(t)$ and $\Lambda_2^G(t)$, respectively ([Andersen et al., 1993](#)), and that $\hat{p}$ is a consistent estimator of p . It is also clear that $\hat{\Lambda}_1(t)$ is independent of $\hat{\Lambda}_2^S(t)$, $\hat{\Lambda}_2^G(t)$ and $\hat{p}$. Furthermore, $\Lambda_2^S(t)$ and $\hat{\Lambda}_2^G(t)$ are independent. We have

$$n_1^{1/2}\{\hat{\Lambda}_1(t) - \Lambda_1(t)\} = n_1^{-1/2} \int_0^t \frac{dM_1(u)}{y_1(u)} + n_1^{-1/2} \int_0^t \left\{ \frac{1}{Y_1(u)/n_1} - \frac{1}{y_1(u)} \right\} dM_1(u). \quad (\text{A1})$$

Assuming $\Lambda_1(\tau) < \infty$, $n_1^{-1/2} M_1$ converges weakly to a zero-mean Gaussian process by the martingale central limit theorem ([Andersen et al., 1993](#)). It follows from [Lin & Ying \(2001, Lemma A.1\)](#) and our Lemma A1 that $n_1^{-1/2} \int_0^t [1/\{Y_1(u)/n_1\} - 1/y_1(u)] dM_1(u) \rightarrow 0$ in probability uniformly in $t \in [0, \tau]$. This, together with (A1), implies that

$$n^{1/2}\{\hat{\Lambda}_1(t) - \Lambda_1(t)\} = n_1^{-1/2} \rho_1^{-1/2} \sum_{i=1}^{n_1} \int_0^t \frac{dM_{1i}(u)}{y_1(u)} + o_p(1) \quad (\text{A2})$$

uniformly in $t \in [0, \tau]$. Similarly, we can show that as $n \rightarrow \infty$ and $n_2/n \rightarrow \rho_2$,

$$n^{1/2}\{\hat{\Lambda}_2^S(t) - \Lambda_2^S(t)\} = n_2^{-1/2} \rho_2^{-1/2} \sum_{i=1}^{n_2} \int_0^t \frac{dM_{2i}^S(u) g_{2i}}{y_2^S(u)} + o_p(1) \quad (\text{A3})$$

and

$$n^{1/2}\{\hat{\Lambda}_2^G(t) - \Lambda_2^G(t)\} = n_2^{-1/2} \rho_2^{-1/2} \sum_{i=1}^{n_2} \int_0^t \frac{dM_{2i}^G(u)(1 - g_{2i})}{y_2^G(u)} + o_p(1) \quad (\text{A4})$$

uniformly in $t \in [0, \tau]$.

Let $\mathcal{D}[0, \tau]^3$ be the metric space consisting of $\{f_1(t), f_2(t), f_3(t)\}$, where $f_k : [0, \tau] \rightarrow R$ for $k = 1, 2, 3$ are right-continuous functions with left limits. The metric of $\mathcal{D}[0, \tau]^3$ is defined as $d(f, g) = \max_{t \in [0, \tau]} \{\|f_k(t) - g_k(t)\| : 1 \leq k \leq 3\}$ for $f, g \in \mathcal{D}[0, \tau]^3$. It is easy to see from equations (A2)–(A4) that the stochastic process $n^{1/2}[\{\hat{\Lambda}_1(t) - \Lambda_1(t)\}, \{\hat{\Lambda}_2^S(t) - \Lambda_2^S(t)\}, \{\hat{\Lambda}_2^G(t) - \Lambda_2^G(t)\}, (\hat{p} - p)]$ in $\mathcal{D}[0, \tau]^3 \times R$ is asymptotically equivalent to a sum of independent and identically distributed random vectors. By the multivariate central limit theorem, its finite-dimensional distributions converge asymptotically to zero-mean multivariate normal distributions. Moreover, because the elements, $n^{1/2}\{\hat{\Lambda}_1(t) - \Lambda_1(t)\}$, $n^{1/2}\{\hat{\Lambda}_2^S(t) - \Lambda_2^S(t)\}$ and $n^{1/2}\{\hat{\Lambda}_2^G(t) - \Lambda_2^G(t)\}$, are square-integrable martingales with respect to their marginal filtrations, their tightness follows from the proof of [Pollard \(1984, Theorem VIII.13\)](#). Hence, $n^{1/2}[\{\hat{\Lambda}_1(t) - \Lambda_1(t)\}, \{\hat{\Lambda}_2^S(t) - \Lambda_2^S(t)\}, \{\hat{\Lambda}_2^G(t) - \Lambda_2^G(t)\}, (\hat{p} - p)]$ converges weakly to a zero-mean Gaussian stochastic process $\{\mathcal{W}_1(t), \mathcal{W}_2(t), \mathcal{W}_3(t), \mathcal{W}_4\}$ in $\mathcal{D}[0, \tau]^3 \times R$ with the variance-covariance functions between $\mathcal{W}_1(t_1)$, $\mathcal{W}_2(t_2)$, $\mathcal{W}_3(t_3)$ and $\mathcal{W}_4$ given by

$$\begin{pmatrix} \rho_1^{-1} \sigma_{11}(t_1) & 0 & 0 & 0 \\ 0 & \rho_2^{-1} \sigma_{22}(t_2) & 0 & \rho_2^{-1} \sigma_{24}(t_2) \\ 0 & 0 & \rho_2^{-1} \sigma_{33}(t_3) & \rho_2^{-1} \sigma_{34}(t_3) \\ 0 & \rho_2^{-1} \sigma_{24}(t_2) & \rho_2^{-1} \sigma_{34}(t_3) & \rho_2^{-1} p(1 - p) \end{pmatrix}.$$

Recall that $\hat{H}_1(t) = \prod_{u \leq t} \{1 - d\hat{\Lambda}_1(u)\}$, $\hat{S}_2(t) = \prod_{u \leq t} \{1 - d\hat{\Lambda}_2^S(u)\}$ and $\hat{G}_2(t) = \prod_{u \leq t} \{1 - d\hat{\Lambda}_2^G(u)\}$, respectively. It follows from the functional delta method ([Andersen et al., 1993](#)) that the joint stochastic process $n^{1/2}[\{\hat{H}_1(t) - H_1(t)\}, \{\hat{S}_2(t) - S_2(t)\}, \{\hat{G}_2(t) - G_2(t)\}, (\hat{p} - p)]$ converges weakly to a zero-mean Gaussian process $\{\mathcal{W}_1^*(t), \mathcal{W}_2^*(t), \mathcal{W}_3^*(t), \mathcal{W}_4^*\}$ with the variance-covariance function between

$\mathcal{W}_1^*(t_1)$, $\mathcal{W}_2^*(t_2)$, $\mathcal{W}_3^*(t_3)$ and $\mathcal{W}_4^*$ given by

$$\begin{pmatrix} H_1^2(t_1)\rho_1^{-1}\sigma_{11}(t_1) & 0 & 0 & 0 \\ 0 & S_2^2(t_2)\rho_2^{-1}\sigma_{22}(t_2) & 0 & -S_2(t_2)\rho_2^{-1}\sigma_{24}(t_2) \\ 0 & 0 & G_2^2(t_3)\rho_2^{-1}\sigma_{33}(t_3) & -G_2(t_3)\rho_2^{-1}\sigma_{34}(t_3) \\ 0 & -S_2(t_2)\rho_2^{-1}\sigma_{24}(t_2) & -G_2(t_3)\rho_2^{-1}\sigma_{34}(t_3) & \rho_2^{-1}p(1-p) \end{pmatrix}.$$

Therefore, for any $t \in [0, \tau]$, $n^{1/2}[\{\hat{H}_1(t) - H_1(t)\}, \{\hat{S}_2(t) - S_2(t)\}, \{\hat{G}_2(t) - G_2(t)\}, (\hat{p} - p)]$ converges to a multivariate normal distribution with mean zero and variance-covariance matrix $\Sigma(t)$.

To prove the consistency of $\hat{\sigma}_{11}(t)$, we have

$$\begin{aligned} \hat{\sigma}_{11}(t) &= \frac{1}{n_1} \sum_{i=1}^{n_1} \left[\int_0^t \left\{ \frac{1}{Y_1(u)/n_1} - \frac{1}{y_1(u)} \right\} d\hat{M}_{1i}(u) \right]^2 \\ &\quad + \frac{2}{n_1} \sum_{i=1}^{n_1} \left[\int_0^t \left\{ \frac{1}{Y_1(u)/n_1} - \frac{1}{y_1(u)} \right\} d\hat{M}_{1i}(u) \int_0^t \frac{d\hat{M}_{1i}(u)}{y_1(u)} \right] + \frac{1}{n_1} \sum_{i=1}^{n_1} \left\{ \int_0^t \frac{d\hat{M}_{1i}(u)}{y_1(u)} \right\}^2. \end{aligned}$$

By the uniform consistency of $\hat{\Lambda}_1(t)$ and $Y_1(t)/n_1$ and the fact that $|Y_{1i}(t)| \leq 1$, we have

$$\begin{aligned} \sup_{t \in [0, \tau]} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} \left[\int_0^t \left\{ \frac{1}{Y_1(u)/n_1} - \frac{1}{y_1(u)} \right\} d\hat{M}_{1i}(u) \right]^2 \right| &\rightarrow 0, \\ \sup_{t \in [0, \tau]} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} \left[\int_0^t \left\{ \frac{1}{Y_1(u)/n_1} - \frac{1}{y_1(u)} \right\} d\hat{M}_{1i}(u) \int_0^t \frac{d\hat{M}_{1i}(u)}{y_1(u)} \right] \right| &\rightarrow 0, \\ \sup_{t \in [0, \tau]} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} \left\{ \int_0^t \frac{d\hat{M}_{1i}(u)}{y_1(u)} \right\}^2 - \frac{1}{n_1} \sum_{i=1}^{n_1} \left\{ \int_0^t \frac{dM_{1i}(u)}{y_1(u)} \right\}^2 \right| &\rightarrow 0, \end{aligned}$$

in probability. Thus, $\hat{\sigma}_{11}(t) = n_1^{-1} \sum_{i=1}^{n_1} \left\{ \int_0^t dM_{1i}(u)/y_1(u) \right\}^2 + o_p(1)$ uniformly in $t \in [0, \tau]$. Furthermore, it follows from the uniform law of large numbers (Newey & McFadden, 1994, Lemma 2.4) that

$$\sup_{t \in [0, \tau]} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} \left\{ \int_0^t \frac{dM_{1i}(u)}{y_1(u)} \right\}^2 - E \left[\left\{ \int_0^t \frac{dM_{1i}(u)}{y_1(u)} \right\}^2 \right] \right| \rightarrow 0$$

in probability. Therefore, $\hat{\sigma}_{11}(t) \rightarrow E[\{\int_0^t dM_{1i}(u)/y_1(u)\}^2] \equiv \sigma_{11}(t)$ in probability uniformly in $t \in [0, \tau]$.

The uniform consistency of $\hat{\sigma}_{22}(t)$, $\hat{\sigma}_{33}(t)$, $\hat{\sigma}_{24}(t)$ and $\hat{\sigma}_{34}(t)$ can be proved similarly. The consistency of $\hat{p}(1 - \hat{p})$ follows directly from the weak law of large numbers. $\square$

Proof of Theorem 1. It follows from Lemma 1 and the delta method that at a given time-point $t \in [0, \tau]$, $n^{1/2}[\{\hat{S}_1(t) - S_1(t)\}, \{\hat{S}_2(t) - S_2(t)\}]$ is asymptotically normally distributed with mean zero and variance-covariance matrix $f\Sigma(t)f'$, where

$$f = \begin{pmatrix} 1/p & 0 & -(1-p)/p & -\{H_1(t) - G_2(t)\}/p^2 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Applying Slutsky's theorem, we have $\hat{v}_{11}(t) \rightarrow v_{11}(t)$, $\hat{v}_{22}(t) \rightarrow v_{22}(t)$ and $\hat{v}_{12}(t) \rightarrow v_{12}(t)$ in probability as $n \rightarrow \infty$, $n_1/n \rightarrow \rho_1$ and $n_2/n \rightarrow \rho_2$. $\square$

Proof of Theorem 2. The theorem can be proved using essentially the same arguments given in the appendix of Berger & Boos (1999). Thus, we omit the details. $\square$

REFERENCES

- ALBERT, J. M. (2002). Estimating efficacy in clinical trials with clustered binary responses. *Statist. Med.* **21**, 649–61.
- ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.
- BERGER, R. L. & BOOS, D. D. (1999). Confidence limits for the onset and duration of treatment effect. *Biomet. J.* **41**, 517–31.
- CHUNG, K. L. (2001). *A Course in Probability Theory*, 3rd ed. New York: Academic Press.
- CUZICK, J., EDWARDS, R. & SEGNA, N. (1997). Adjusting for non-compliance and contamination in randomized clinical trials. *Statist. Med.* **16**, 1017–29.
- CUZICK, J., SASIENI, P., MYLES, J. & TYRER, J. (2007). Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *J. R. Statist. Soc. B* **69**, 565–88.
- FRANGAKIS, C. E. & RUBIN, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-non-compliance and subsequent missing outcomes. *Biometrika* **86**, 365–79.
- GOETHGHEBEUR, E. & MOLENBERGHS, G. (1996). Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *J. Am. Statist. Assoc.* **91**, 928–34.
- HOLLIS, S. & CAMPBELL, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *Br. Med. J.* **319**, 670–4.
- KAPLAN, E. L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.* **53**, 457–81.
- KEMENY, M. M., ADAK, S., GRAY, B., MACDONALD, J. S., SMITH, T., LIPSITZ, S., SIGURDSON, E. R., O'DWYER, P. J. & BENSON, A. B. (2002). Combined-modality treatment for resectable metastatic colorectal carcinoma to the liver: surgical resection of hepatic metastases in combination with continuous infusion of chemotherapy—an intergroup study. *J. Clin. Oncol.* **20**, 1499–505.
- LACHIN, J. M. (2000). Statistical considerations in the intent-to-treat principle. *Contr. Clin. Trials* **21**, 167–89.
- LIN, D. Y. & YING, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Am. Statist. Assoc.* **96**, 103–13.
- LOEYS, T. & GOETGHEBEUR, E. (2003). A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics* **59**, 100–5.
- LOEYS, T., GOETGHEBEUR, E. & VABDEBOSCH, A. (2005). Casual proportional hazards models and time-consistent exposure in randomized clinical trials. *Lifetime Data Anal.* **11**, 435–49.
- MORTON, D. L., THOMPSON, J. F., COCHRAN, A. J., MOZZILLO, N., ELASHOFF, R., ESSNER, R., NIEWEG, O. E., ROSES, D. F., HOEKSTRA, H. J., KARAKOUSIS, C. P., ET AL. (2006). Sentinel-node biopsy or nodal observation in melanoma. *New Engl. J. Med.* **355**, 1307–17.
- NEWKEY, W. K. & MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, vol. 4, Ch. 36, pp. 2111–245. Amsterdam: Elsevier Science.
- POLLARD, D. (1984). *Weak Convergence of Stochastic Processes*. New York: Springer.
- ROBINS, J. M. & TSIATIS, A. A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Commun. Statist. A* **20**, 2609–31.
- SOMMER, A. & ZEGGER, S. L. (1991). On estimating efficacy from clinical trials. *Statist. Med.* **10**, 45–53.
- ZELEN, M. (1979). A new design for randomized clinical trials. *New Engl. J. Med.* **300**, 1242–5.
- ZELEN, M. (1990). Randomized consent designs for clinical trials: an update. *Statist. Med.* **9**, 645–56.

[Received February 2010. Revised November 2011]