A sensitivity analysis for causal parameters in structural proportional hazards models

E. Goetghebeur* and T. Loeys

Ghent University

Abstract

Deviations from assigned treatment occur often in clinical trials. In such a setting, the traditional intent-to-treat analysis does not measure biological efficacy but rather programmatic effectiveness. For all-or-nothing compliance situation, Loeys and Goetghebeur (2003) recently proposed a Structural Proportional Hazards method. It allows for causal estimation in the complier subpopulation provided the exclusion restriction holds: randomization per se has no effect unless exposure has changed. This assumption is typically made with structural models for noncompliance but questioned when the trial is not blinded. In this paper we extend the structural PH model to allow for an effect of randomization per se. This enables analyzing sensitivity of conclusions to deviations from the exclusion restriction. In a colo-rectal cancer trial we find the causal estimator of the effect of an arterial device implantation to be remarkably insensitive to such deviations.

1 Introduction

While the randomized clinical trial remains the gold standard design for causal inference, a thorough analysis of the impact of an intervention should consider treatment actually received besides treatment assigned. The distance between intended and materialized treatments can indeed vary widely, first inside the trial and later under less controlled

conditions. Hence the challenge to estimate the effect of different levels of treatment that occur in practice.

The hazard ratio has become the most popular measure of the effect of treatment on survival. Intention-to-treat results are typically cast in those terms, the theory has been well developed and resulting estimators well understood. On the other hand, structural accelerated failure time models (SAFT) as proposed by Robins and Tsiatis (1991) have become the usual tools for 'causal survival analysis' conditional on observed exposures. These models express how survival time can be shrunk or expanded by a parametric function of observed exposures to yield potential treatment-free survival times. In the absence of a direct effect of randomization, potential treatment-free survival times are by design equally distributed between randomized arms. To tap into the proportional hazards tradition and allow for a smooth exchange of information, Loeys and Goetghebeur (2003) developed structural proportional hazards models. They analyzed ECOG-trial E9288 (Kemeny *et al.*, 2002), a randomized clinical trial in colorectal cancer patients with liver metastases. This ECOG multi-centre trial was initiated because the long-term outcome of resection of hepatic metastases remained poor and arterial chemotherapy regimens targeted to the liver had demonstrated high potential. Patients were randomly assigned to either surgical resection alone (control arm, 56 patients) or surgical resection followed by chemotherapy (experimental arm, 53 patients). Interest focused on comparing 5-year survival with and without the implantation. The multi-centric nature of the study made preoperative randomization the practical option. As a result, ten patients who were randomized to receive experimental treatment, did not receive the arterial device implantation, possibly for reasons related to their survival chance.

At the First Barcelona Workshop on Survival Analysis, Ross Prentice raised questions concerning the exclusion restriction given the unblinded nature of the study. It is not inconceivable for instance that the bad news of not being able to receive the implant once it was planned had a negative effect on the patients outcome. Likewise, surgery involving an intended implant might be scheduled earlier in the day, which may have its own impact on survival etc. In response to such concerns this paper sets out to conduct a sensitivity analysis as follows.

In Section 2 we provide a rationale for causal methodology in a proportional hazards framework. Section 3 details the structural proportional hazards approach under the exclusion restriction. In Section 4 we extend the model and adapt the estimation procedure to allow for a sensitivity analysis and examine the impact of violations of the exclusion restriction on the causal PH-estimator. In Section 5 we move on to investigate the joint effect of assignment in both the noncompliant ('the exclusion effect') and compliant ('the causal effect') subpopulation. The methodology is applied to the E9288-data in Section 6 and discussed in Section 7.

2 Rationale for a causal proportional hazards estimator

When clinical trial participants fail to adhere to their assigned treatment, a straightforward but naive estimator of the effect of treatment actually received compares patients who were observed to receive the experimental exposure with those who did not. Consider specifically the Cox model

$$h(t \mid E_i) = h_0(t) \exp(\beta_0 E_i), \tag{1}$$

where E_i is the all-or-nothing exposure indicator and $h(t \mid E_i)$ is the hazard rate for failure at time t given exposure. When compliance is selective, i.e. when individuals who comply are prognostically different from those who do not, the parameter β_0 carries no causal interpretation.

Therefore the most commonly used approach is an intent-to-treat analysis. In proportional hazard terms:

$$h(t \mid R_i) = h_0(t) \exp(\gamma_0 R_i), \tag{2}$$

where $R_i = 1$ indicates the experimental arm. The advantage of this approach is its validity under the null. When experimental treatment has no effect, survival distributions coincide on both randomized arms and $\gamma_0 = 0$ corresponds to the true model. However, in the presence of non-compliance, γ_0 does not generally measure the biological effect of treatment but rather mixes the effect on compliers with the absence of effect on non-compliers.

To estimate the causal effect of treatment actually received, structural models can be used. Loeys and Goetghebeur (2003) consider

$$h(t \mid R_i = 1, U_i = u) = h(t \mid R_i = 0, U_i = u) \exp(\psi_0 u) \tag{3}$$

where U_i is the potential all-or-nothing exposure for the ith subject, that is the exposure that would have been observed had subject i been randomized to experimental treatment. U_i is observed on the experimental arm but latent on the control arm. The Causal Proportional Hazards Effect of Treatment (C-PROPHET) is the log hazard ratio ψ_0 in model (3). It compares survival under experimental and potential control conditions in the treatable subgroup $\{U_i = 1\}$. A negative (respectively positive) ψ_0 implies a beneficial (respectively harmful) effect of implantation in the treatable subset.

In the subgroup $\{U_i = 0\}$ that would not have been treated when assigned to experimental treatment, no effect of assignment on survival is assumed. Imbens and Rubin (1997) call this assumption the 'exclusion restriction', while Pearl (2002) calls this 'the absence of indirect effect'. The main challenge for inference in model (3) stems from U_i being unobserved in the control arm. We summarize in the next section how ψ_0 can be estimated under the exclusion restriction despite ignoring this potential compliance information.

3 Inference for the C-PROPHET estimator

If potential receivers of the experimental exposure were known at baseline in both arms, one would fit a proportional hazards model in the subgroup $\{U_i = 1\}$. Denote then Breslow's cumulative baseline hazard estimator for survival in the $\{R_i = 0, U_i = 1\}$-group by $\hat{H}_{01}(t)$. Within the $\{U_i = 1\}$-subset the partial likelihood score equation can be rewritten - in the absence of ties - as

$$\sum_{t_{(j)}} \left[R_{(j)} - \left\{ \hat{H}_{01}(t_{(j)}) - \hat{H}_{01}(t_{(j-1)}) \right\} n_{11j} e^{\psi_0} \right] = 0 \tag{4}$$

where $t_{(j)}$ is the j-th ordered failure time in the treatable subset and $R_{(j)}$ the corresponding assignment indicator. It thus suffices to estimate the jumps of the cumulative hazard for the unobserved subset of compliers in the control arm. H_{01} can be estimated via the corresponding survival estimator $\hat{S}_{01}^*(t)$,

$$\widehat{S}_{01}^*(t) = \{\widehat{S}_0(t) - (1 - \widehat{\pi})\widehat{S}_{10}(t)\}/\widehat{\pi}, \tag{5}$$

with $S_r(t) := \Pr(T_i > t \mid R_i = r)$, $S_{ru}(t) := \Pr(T_i > t \mid R_i = r, U_i = u)$, and $\widehat{\pi}$ the observed compliance proportion in the experimental arm. The Kaplan-Meier estimates $\widehat{S}_0(t)$ and $\widehat{S}_{10}(t)$ will be consistent under independent censoring or the weaker assumption that censoring is non-informative for the control arm as a whole, while in the experimental arm censoring is non-informative conditional on treatment exposure. Frangakis and Rubin (1999) argue that it is sometimes more reasonable to assume non-informative censoring on potential treatment exposure in both arms, and showed that even under this scenario $S_{01}(t)$ is identifiable. Because $\widehat{S}_{01}^*(t)$ is not necessarily monotonic decreasing and found to be a poor estimator for $S_{01}(t)$, Loeys and Goetghebeur (2002) suggested to improve on the proposed estimator via isotonic regression and the 'Pool-Adjacent-Violators' Algorithm (Barlow *et al.*, 1972). To avoid ties in bootstrap samples (Efron, 1981), the jackknife procedure is proposed for variance estimation. Simulation revealed that this procedure provides a somewhat conservative variance estimator in this setting.

4 The exclusion restriction: a sensitivity analysis

The exclusion restriction implied by model (3) disallows an effect of assignment for (potential) non-receivers. While this is plausible in double-blind settings, our motivating example E9288 was unblinded.

Consider therefore the following pair of causal models:

$$\begin{cases} h(t \mid R_i = 1, U_i = 0) = h(t \mid R_i = 0, U_i = 0) \exp(\eta_0) \\ h(t \mid R_i = 1, U_i = 1) = h(t \mid R_i = 0, U_i = 1) \exp(\psi_0) \end{cases} \tag{6}$$

In the treatable subset $\{U_i = 1\}$ we consider a proportional hazards effect of exposure as before, but in the $\{U_i = 0\}$-subset we no longer require equality in distribution

between randomized arms. Instead, a positive η_0 in (6) implies that omission of an implantion that was assigned is bad news added to bad news, i.e. observing the inability of an implant on the experimental arm deteriorates the already bad survival prognosis compared to not observing this on the control arm. This additionally imposed proportional hazard assumption in the $\{U_i = 0\}$-subset can be rewritten in terms of the survival distributions:

$$S_{00}(t) = S_{10}(t)^{\exp(-\eta_0)}. \tag{7}$$

Using equality (7), we obtain a treatment-free survival curve for potential compliers from

$$\widehat{S}_{01}^{*}(t;\eta_0) = \{\widehat{S}_0(t) - (1 - \widehat{\pi})\widehat{S}_{10}(t)^{\exp(-\eta_0)}\}/\widehat{\pi}. \tag{8}$$

As before, the pointwise estimator need not be monotone and isotonic regression on time yields our estimator $\hat{S}_{01}(t;\eta_0)$. Upon substituting the monotonized $\hat{S}_{01}$ to obtain $\tilde{H}_{01}(t;\eta_0) = -\log\widehat{S}_{01}(t;\eta_0)$, we estimate ψ_0 in function of η_0 as

$$\exp(\hat{\psi}_0(\eta_0)) = \frac{\sum_{t_{(j)}} R_{(j)}}{\sum_{t_{(j)}} \{\tilde{H}_{01}(t_{(j)};\eta_0) - \tilde{H}_{01}(t_{(j-1)};\eta_0)\} n_{11j}}. \tag{9}$$

5 Joint estimation of effect in the compliant and noncompliant subpopulation

In the previous section ψ_0 is estimated as a function of a fixed sensitivity parameter η_0. Next we investigate joint estimation of η_0 and ψ_0.

The estimation procedure outlined in Section 3 is restricted to all-or-nothing compliance. Indeed, when several levels of compliance are involved, the survival distribution $S_{01}(t)$ is no longer identified without strong additional assumptions. Recently, Loeys and Goetghebeur (2002) proposed an alternative estimation procedure that overcomes this limitation. Specifically, they backtransform observed survival distributions in the experimental arm by exponential functions of the measured exposures to obtain treatment-free survival distributions. Averaging these over all complier subgroups yields an unconditional treatment-free survival curve in the treatment arm. Under the exclusion restriction this should match the corresponding curve on the control arm. Allowing now for an effect of assignment in the $\{U_i = 0\}$-group as in model (6), the idea is to check whether the distribution of observed survival times in the control arm is close to the new mixture of backtransformed survival distributions observed in the experimental arm:

$$S_{1 \longrightarrow 0}(t;\eta,\psi) = \hat{S}_{10}(t)^{\exp(-\eta)}(1 - \hat{\pi}) + \hat{S}_{11}(t)^{\exp(-\psi)}\hat{\pi}. \tag{10}$$

Parameter values η and ψ which 'equalize' the treatment-free survival distributions between randomized arms are point estimators for η_0 and ψ_0. Because we are estimating two parameters here, two estimating equations are needed. Loeys and Goetghebeur

(2002) combine a logrank and weighted logrank test statistic which are built as sums of 'pseudo' martingale residuals. As the test statistic $Q(\eta,\psi)$ is approximately $\chi^2(2)$, a 95% confidence region for (η_0,ψ_0) is formed by the set of (η,ψ)-values for which $Q(\eta,\psi)$ is below 6.0.

In practice the information may be weak and identification of both η_0 and ψ_0 over-ambitious. Small scale simulations confirm that with limited selectivity (i.e. receivers and non-receivers having a comparable baseline survival prognosis) identifiability problems indeed occur. However with an increasing selection effect both 'causal' effects were reasonably well identified in the simulation setting. Results on the dataset E9288 are described next.
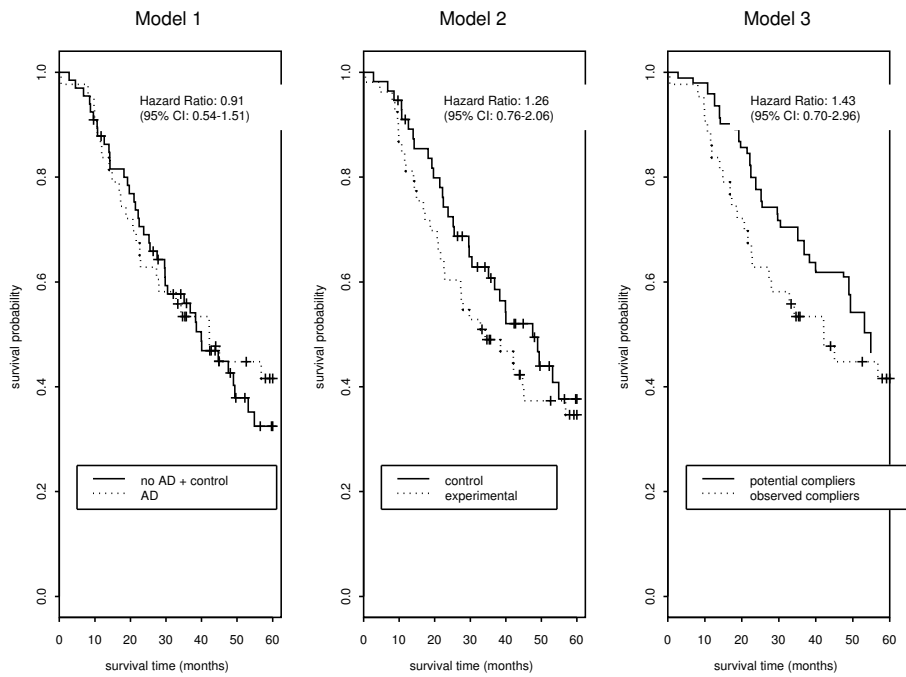

Figure 1: *3 approaches: (1) As treated (2) Intent-to-treat (3) C-PROPHET.*
Note: AD stands for Arterial Device).

6 The causal effect of an arterial device implantation

In this section we analyze the E9288-data. Within the 5-year follow-up, 30 patients (54%) died on the control arm compared to 33 patients (62%) on the experimental arm. Figure 1 shows results following models (1), (2) and (3). The as-treated analysis estimates a non-significant beneficial effect of implantation but is only valid under the assumption that non-receivers on the experimental arm form a random subset from

the entire population. The estimated effect under the intent-to-treat analysis reveals a non-significant harmful effect of assignment, but does not capture the effect of the implantation actually received. The structural analysis reveals a 43% increase in hazard associated with arterial device implantation in the treatable subset. The latter is derived under the exclusion restriction. Under this assumption, it is further shown in Loeys and Goetghebeur (2003) how the selectivity of patients getting the intervention can be presented by contrasting $\widehat{S}_{01}$, as estimated in (5) with $\widehat{S}_{10} = \widehat{S}_{00}$, as observed in non-compliers in the experimental arm. That plot revealed that patients who would not have received the intervention when assigned to it have a much worse intervention-free survival prognosis than patients that would have received the intervention. This selection effect is also seen in the as-treated estimator, which mixes the causal treatment effect in the treatable subgroup with a diluted selectivity effect, and thus differs from the C-Prophet estimator.

Following Section 4 we can now investigate how sensitive the C-PROPHET estimator is to violations of the exclusion restriction. From Figure 2, we learn that under the assumption of a 50% decrease (respectively increase) in hazard associated with implantation assignment in the untreatable subset, the estimated causal effect equals 1.53 with 95% confidence interval ranging from 0.72 to 3.25 (respectively 1.36, with 95% CI: $0.68 - 2.77$). We thus observe that quite substantial deviations from the exclusion restriction have rather limited impact on the estimated causal hazard ratio. Relative to the width of the 95% confidence interval, the change in causal effect as a function of the sensitivity parameter is indeed quite small.
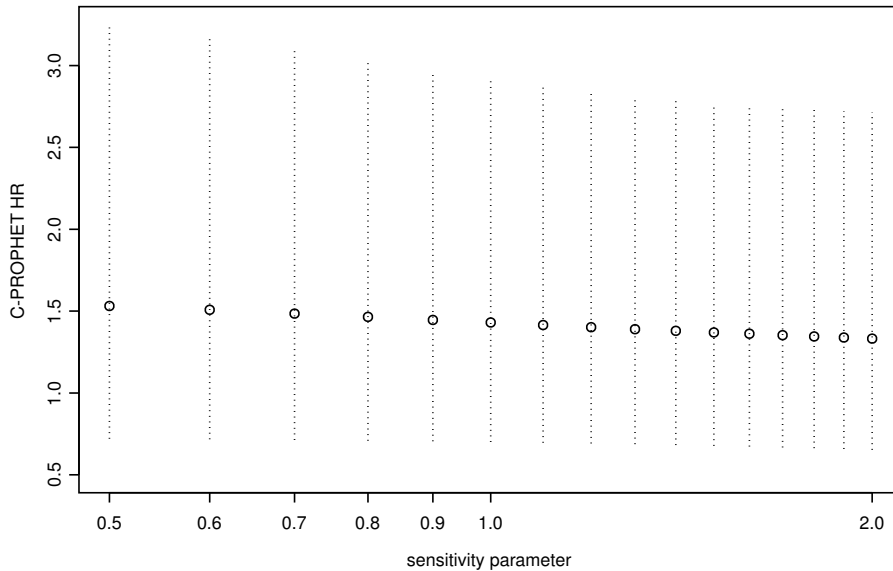

Figure 2: *The causal hazard ratio* $\exp(\psi_0)$ *(with 95% confidence interval) as a function of the sensitivity parameter* $\exp(\eta_0)$.
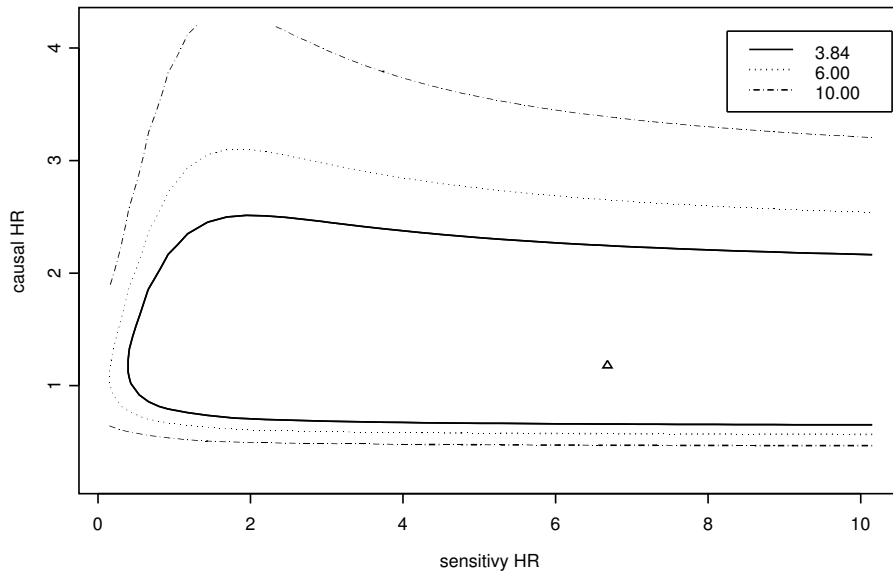
Figure 3: *Joint estimation of the sensitivity hazard ratio* $\exp(\eta_0)$ *and causal hazard ratio* $\exp(\psi_0)$.

Results of joint estimation of η_0 and ψ_0 as outlined in Section 5 are summarized in Figure 3. Contour plots show the value of the $\chi^2(2)$-test statistic $Q(\eta,\psi)$ as a function of $\exp(\eta)$ and $\exp(\psi)$. Point estimators for $\exp(\eta_0)$ and $\exp(\psi_0)$ equal 6.63 and 1.19. As the 95% confidence region does not close for increasing values of η, identifiability of η_0 is rather poor, in contrast to that of ψ_0. Nevertheless the data appear to favour the region of η suggesting that bad news add more bad news. A substantial positive effect of an intended but absent implant is excluded as a possibility. Surprisingly however we see negligeable impact on estimation of the primary parameter. A marginal 95% confidence interval for η_0 and ψ_0 can be found upon projecting the 3.84-contour on the axes (Robins and Greenland, 1994).

7 Discussion

In this paper we presented two approaches to investigate violations against the exclusion restriction in a causal proportional hazards framework. While the approach presented in Section 4 is limited to all-or-nothing compliance, the approach of Section 5 allows for several compliance levels. As identifiability under the second approach relies on the unobserved selectivity, one should be careful when interpreting its results. In a missing data setting Scharfstein (2002) recently discovered that when the signal for inference on η_0 is weak, it is dangerous to believe the η-estimate. He therefore favors a sensitivity approach as proposed in Section 4. Interestingly, Scharfstein (2002) found

narrow confidence intervals conditional on η relative to an enormous range of point estimates for ψ as η varies. In contrast we obtain in our motivating example, despite a non-negligeable portion of observed non-compliers on the experimental arm (10 out of 53), changes in the outcome distribution of potential non-receivers with a relatively small impact on the causal parameter estimates.

Further research would be welcomed on the role baseline predictors for treatment-free survival and/or exposure. In as treated model (1) we cannot expect to capture the dependence between treatment-free survival and potential experimental compliance by conditioning on these baseline covariates, and the as treated approach will still give biased results. Loeys and Goetghebeur (2002) propose an estimation procedure allowing to identify population-averaged Causal PROPortional Hazards Effects of Treatment at observed exposure and covariate levels. Conditioning on baseline covariates in model (3) can then address confounding (in the presence of imbalance between randomized arms), conservatism and/or help to keep censoring non-informative.

Finally it is worth remembering that our approach studies survival models conditional on exposure status. Marginal proportional hazards models have recently been introduced by Hernan, Brumback and Robins (2000).

Acknowledgements

References

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions.* New York: Wiley.

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, 76, 312-319.

Frangakis, C. and Rubin, D. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment noncompliance and subsequent missing outcomes. *Biometrika*, 86, 365-379.

Imbens, G. and Rubin, D. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, 25, 305-327.

Hernßn, M.A., Brumback, B. and Robins, J.M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5), 561-570.

Kemeny, M.M., Adak, S., Gray, B., Macdonald, J.S., Smith, T., Lipsitz, S., Sigurdson, E.R., O'Dwyer, P.J. and 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. *Journal of Clinical Oncology*, 20, 1499-1505.

Loeys, T. and 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, 100-105.

Loeys, T. and Goetghebeur, E. (2002). Causal proportional hazards models for the effect of treatment actually received in a randomized trial with selective noncompliance. *Technical Report Centrum voor Statistiek, Gent*.

Pearl, J. (2002). Causal inference in the health sciences: a conceptual introduction. In *Health Ser. Outcomes Res. Method* (in press).

Robins, J.M. and Tsiatis, A.A. (1991). Correcting for noncompliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics A*, 20 (8), 2609-2631.

Robins, K.M. and Greenland, S. (1994). Adjusting for differential rates of prophylaxis for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association*, 89, 737-749.

Scharfstein, D. (2002). Frequentist and bayesian inference for potentially non-ignorable non-response in randomized clinical trials. *Proceedings IBC 2002*.

Sun, J. and Kalbfleisch, J.D. (1993). The analysis of current status data on point processes. *Journal of the American Statistical Association*, 88, 1449-1455.

Resum

Les desviacions del tractament assignat són comuns en assajos clínics. En aquest context, l'anàlisi tradicional per intenció de tractar no mesura l'eficàcia biològica sinó l'eficàcia de la programació. Per a aquelles situacions on el compliment és total o nul, Loeys and Goetghebeur (2003) proposen un mètode estructural de riscos proporcionals. Aquest mètode permet l'estimació causal en la subpoblació que compleix, sempre que es verifiqui la restricció d'exclusió: l'aleatorització per se no te cap efecte llevat que es canviï l'exposició. Aquesta premissa s'usa en general amb models estructurals per a incompliment però es qüestiona quan l'assaig no és cec. En aquest treball estenem el model estructural de riscos proporcionals de manera que admeti un efecte d'aleatorizació per se. Això ha de permetre analitzar la sensibilitat de les conclusions a les desviacions de la restricció d'exclusió. A un assaig clínic sobre càncer colo-rectal trobem que l'estimador causal de l'efecte de la implantació d'un dispositiu arterial és remarcablement insensible a aquestes desviacions.

MSC: 62N01, 62N02, 62P10, 92B15

Paraules clau: Anàlisi de sensibilitat; assajos clínics aleatoritzats; compliment; inferència causal; restricció d'exclusió; riscos proporcionals