

CORRECTING FOR NON-COMPLIANCE IN RANDOMIZED TRIALS: AN APPLICATION TO THE ATBC STUDY

PASI A. KORHONEN^{1*}, NAN M. LAIRD² AND JUNI PALMGREN³

¹*Rolf Nevanlinna Institute, P.O. Box 4, FIN-00014, Helsinki, Finland*

²*Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, U.S.A.*

³*Mathematical Statistics, Stockholm University, S-10691 Stockholm, Sweden*

SUMMARY

Different methods for estimating the effect of treatment actually received in a longitudinal placebo-controlled trial with non-compliance are discussed. Total mortality from the ATBC Study is used as an illustrative example. In the ATBC Study some 25 per cent of the participants dropped out from active follow-up prior to the scheduled end of the study. The ‘intention-to-treat’ analysis showed an increased death risk in the beta-carotene arm when compared with the no beta-carotene arm. Owing to considerable non-compliance it is also of interest to estimate the effect of beta-carotene actually received. We use a simple model for the treatment action and discuss three methods for estimation of the treatment effect under the model – the ‘intention-to-treat’ approach, the ‘as-treated’ approach and the g-estimation approach. These approaches are compared in a simulation study under different settings for non-compliance. Finally, the data from the ATBC Study are analysed using the proposed methods. Copyright © 1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

In population based randomized chemoprevention studies, complete endpoint information on death is obtained from national registers for all participants. The primary analysis is based on the ‘intention-to-treat’ (ITT) comparison, that is, survival comparison of the treatment groups as randomized, regardless of subsequent adherence to the treatment protocol. By virtue of randomization the ITT comparison provides a valid test for the sharp causal null hypothesis of no treatment effect. However, prior to the scheduled end of follow-up a substantial number of

* Correspondence to: Pasi A. Korhonen, Rolf Nevanlinna Institute, P.O. Box 4, FIN-00014, Helsinki, Finland. E-mail: Pasi.Korhonen@RNI.Helsinki.fi

Contract/grant sponsor: Academy of Finland
Contract/grant numbers: 31116, 33495

Contract/grant sponsor: Yrjö Jahnsson Foundation
Contract/grant numbers: 3498, 3816

Contract/grant sponsor: NIH
Contract/grant number: GM-29745

Contract/grant sponsor: National Cancer Institute
Contract/grant number: NOI-CN-45165

CCC 0277–6715/99/212879–19\$17.50
Copyright © 1999 John Wiley & Sons, Ltd.

Received October 1997
Accepted February 1999

participants 'drop out', that is, they cease to attend clinic visits and thus cease to receive supplementation. Assessment of the magnitude of the effect of 'treatment actually received' is complicated by the fact that those who drop out prematurely may differ from compliers on important time constant and time varying risk factors, some of which may be unobserved. We assume that various demographic and life-style characteristics known to be predictive of both treatment-free survival and the time on active treatment are measured at baseline, but we do not assume that the full set of such factors has been captured. Here we restrict attention to time constant baseline covariates and do not deal with observed or unobserved time varying confounders.

Our interest stems from the Alpha-Tocopherol Beta Carotene Lung Cancer Prevention Study (ATBC Study), a randomized placebo controlled prevention trial designed to assess the effect of alpha-tocopherol and beta-carotene supplementation on lung cancer incidence in a cohort of 29,113 smoking middle aged men in Finland.¹ The ITT analysis suggests a harmful effect of beta-carotene supplementation on lung cancer incidence and mortality.² Since about 25 per cent withdrew from active treatment prematurely for reasons other than death it is of interest also to assess the effect of beta-carotene actually received. We introduce for each individual the concept of treatment-free survival time which is a potential outcome. We assume that each day on active treatment induces an independent increment Δ to the survival time. For positive Δ the survival time is prolonged (protective effect) and for negative Δ it is contracted (harmful effect). The data observed for a given individual are determined by the treatment arm assignment and by censoring due to the scheduled end of follow-up. It is the structural parameter Δ that is of interest, and we discuss complications that may arise as well as assumptions that need to be made to identify Δ from the observed data.

The treatment-free survival time differs between individuals and can be viewed as a random intercept or error term in the structural model. Initially we make the rather artificial assumption that the potential treatment-free survival time and the potential time on active treatment vary independently of each other in the cohort. We introduce a transformation of the potential variables to observed variables and we describe under what conditions a multiplicative hazards model with a time dependent indicator for the active treatment captures the structural parameter Δ . We proceed by replacing the assumption of independence by the complication of outcome dependent drop-out. We attempt to retrieve the structural parameter Δ using three alternative approaches. (i) ITT analysis; (ii) 'as-treated' (AT) analysis; and (iii) so-called g-estimation (GE analysis). Since our focus is on estimating the magnitude of the effect of treatment actually received on survival, the ITT approach is known to give estimates that are biased towards the null when non-compliance is present. We show that the AT analysis gives valid estimates if the potential time on active treatment is independent of treatment-free survival, and if the treatment-free survival time follows an exponential distribution. Both the ITT and AT methods fail if the potential time on active treatment depends on the treatment-free survival time. We can estimate the parameter in the structural model using g-estimation, which has connections to the use of instrumental variables in econometrics,³ and was originally introduced by Robins.⁴ It draws on the independence between the potential treatment-free survival and the treatment arm allocation. Here we expand on work by Mark and Robins⁵ by explicitly considering the performance of the different methods under outcome dependent non-compliance.

In Section 2 we briefly present the ATBC Study. In Section 3 we describe the observed data and the structural model together with the central assumptions on how treatment affects outcome. In Sections 4, 5 and 6, respectively, we present the ITT, AT and GE methods including their strengths and weaknesses in capturing the structural model parameter. In Section 7 we present

a simulation study under different compliance settings. In Section 8 we analyse the ATBC Study data on beta-carotene supplementation and total mortality and in Section 9 we discuss how the proposed methods relate to the existing literature and how the simple deterministic structural model could be extended to more realistic situations involving differential treatment action.

2. THE ATBC STUDY

The rationale, design, measures of compliance and characteristics of the participants of the alpha-tocopherol beta-carotene (ATBC) study have been described in detail elsewhere.¹ Briefly, this was a randomized double-blinded placebo-controlled chemoprevention trial conducted in 14 adjoining areas in Finland between 1985 and 1993. The primary objective was to evaluate the effects of alpha-tocopherol and beta-carotene supplementations on lung cancer incidence in a cohort of 29,133 smoking middle aged men at high risk for lung cancer. Secondary objectives were to evaluate the effect of supplementation on the incidence of other major cancers and on total mortality.

Participants were recruited from each of the 14 areas between 1985 and 1988. The date of scheduled end of active intervention was 30 April 1993 for all participants. The median follow-up time was about 6 years. Within each area a 2×2 factorial scheme was used to randomize the participants to one of the four possible treatments, alpha-tocopherol alone, beta-carotene alone, their combination, or placebo. At baseline participants were interviewed to obtain details of demographic variables, as well as medical, smoking, dietary and occupational history. Levels of alpha-tocopherol and beta-carotene in serum were also measured from blood samples drawn at baseline. Follow-up consisted of three annual visits to the local field centre, during which the men were asked about their health status and smoking habits since the last visit. During the follow-up visits the study treatment from the previous period were returned, residual capsules were counted and recorded, and a new pack of supplements with a 4-month supply was given to each participant. If the participants stopped attending follow-up visits they could no longer get the study agents and were considered drop-outs. Endpoint information was received from national registers regardless of the drop-out status. Specifically, cancers were identified primarily via the Finnish Cancer Registry and deaths were confirmed via the Central Population Register.

In this paper we focus on analysing the effect of beta-carotene supplementation on total mortality. We compare the beta-carotene group (BC) with the no beta-carotene (NOBC) group. Note that the BC group includes the combination treatment as well as the beta-carotene treatment alone. Similarly, the NOBC group includes placebo treatment and alpha-tocopherol treatment alone. The ITT analysis suggested a harmful effect of beta-carotene supplementation on total mortality.² Since about 25 per cent withdrew from active treatment prematurely for reasons other than death (Table IV) it is of interest also to assess the effect of beta-carotene actually received. We will contrast the ITT analysis, the AT analysis and the GE analysis using the ATBC study as an example in Section 8.

3. DATA AND THE MODEL

3.1. Observed data

We present a general notation for a randomized clinical trial with non-compliance. We assume staggered entry into the trial and a fixed closing date. Each individual is followed from the time of

randomization to death or to the closing date, whichever comes first. Thus the censoring time is fixed for each individual before allocation to the treatment arm.

For subject i we observe a set of baseline covariates, Z_i , which include age, smoking history and other physiological or life-style factors assumed related to the endpoint. The treatment assignment indicator is denoted by R_i and takes the value 0 if allocated to placebo and 1 if allocated to active treatment. T_i is the time to death and C_i is the censoring time, that is, time from randomization to the closing date. We can only observe $X_i = \min(T_i, C_i)$ with $\delta_i = 1$ if $X_i = T_i$, that is, if death occurs before the closing date, and $\delta_i = 0$ otherwise. The time on active treatment is denoted by D_i . In the placebo arm $D_i = 0$, for all i , implying that the treatment agent is not available outside the study setting. In the active treatment arm $D_i \leq X_i$ and we call subject i a complier if $D_i = X_i$ and a non-complier otherwise. We assume monotone drop-out, that is, once a subject has dropped out he cannot re-enter to receive the active treatment. Note, that drop-out in the placebo arm is ignored, since it is not assumed to carry information relevant for compliance in the treated arm. Thus for subject i the observed data are $\{Z_i, R_i, X_i, \delta_i, D_i\}$.

3.2. The structural model

For subject i we define U_i as the potential survival time had no treatment been received. In the ATBC Study U_i denotes the survival time if no beta-carotene was received. For a given time on active treatment D_i we link the treatment-free survival time U_i to survival time T_i through a structural model

$$T_i = U_i + \Delta D_i \quad (1)$$

with $D_i \leq T_i$ and $-\infty < \Delta < 1$ and where D_i and U_i are possibly correlated. Model (1) quantifies how treatment-free survival time is extended or contracted if subject i spends D_i time units on active treatment. Note that U_i is defined for each individual at the time of randomization and is treated as a potential outcome that is fixed at baseline but only partially observed. In the absence of censoring we observe $T_i = U_i$ in the placebo arm because $D_i = 0$, but in the treatment arm U_i is observed only if $D_i = 0$.

Model (1) corresponds to an accelerated failure time model (AFT) with a time varying covariate process $A_i(t) = 1$ if subject i is on active treatment at time t and $A_i(t) = 0$ otherwise. For a fixed Δ we write

$$U_i(\Delta) = \int_0^{T_i} \exp(\psi A_i(s)) ds = \int_0^{D_i} \exp(\psi) ds + \int_{D_i}^{T_i} ds = D_i \exp(\psi) + (T_i - D_i) = T_i - \Delta D_i \quad (2)$$

with $-\infty < \psi < \infty$ and $-\infty < \Delta = 1 - \exp(\psi) < 1$. Robins and Tsiatis⁶ refer to this model as a rank preserving structural nested failure time (RPSNFT) model and they interpret $1/(1 - \Delta)$ as the fractional increase or decrease in survival time if subject i were always on active treatment as opposed to never being on active treatment. Thus if $\Delta = 0.5$ the remaining lifetime is doubled if always on active treatment and if $\Delta = -1$ it is halved if always treated as compared to never treated.

3.3. Assumptions

We make four assumptions, which are in line with those for the instrumental variables approach in linear models where the error process is correlated with the covariate process.³

Assumption 1. Model (1) correctly captures the biological treatment action.

Assumption 2. Randomization takes place at baseline and the assignment mechanism is known. The simplest case involves $\Pr(R_i = 1) = \Pr(R_i = 0) = \frac{1}{2}$.

Assumption 3. The treatment-free survival time U_i for an individual i is unaffected by the treatment assignment or by the survival experience of other individuals.

Assumption 4. The exclusion restriction holds, that is, no other factor than the amount of the treatment received D_i induces a difference in survival experience for an individual i under different treatment arm assignments. This assumption is plausible if the study is appropriately blinded.

In what follows we discuss three methods for estimating the structural parameter in the model (1).

4. THE 'INTENTION-TO-TREAT' (ITT) ANALYSIS

In the 'intention-to-treat' analysis the treatment groups are compared as assigned regardless of non-compliance. The simplest standard analysis for assessing treatment effect in randomized trials with a survival endpoint, possibly non-informatively right censored, is to assume that the proportional hazards assumption holds, that is

$$\lambda_i(t) = \lambda_0(t) \exp(\psi_H R_i) \quad (3)$$

where $\lambda_0(t)$ is the baseline hazard and $\exp(\psi_H)$ the risk ratio parameter. Estimation of the parameter ψ_H based on the partial likelihood was introduced by Cox⁷ and large sample properties of the estimate were justified using martingale theory by Andersen and Gill.⁸ The logrank test, which is equivalent to the partial likelihood score test, can be used for testing of the sharp causal null hypothesis of no treatment effect, that is, $H_0: \psi_H = 0$.

If all subjects adhere to their original treatment assignment throughout the study, that is, $A_i(t) = R_i$ for all t , then from (2) $T_i = U_i \exp(-\psi R_i)$ and after taking logarithms of both sides we get the ITT accelerated failure time model

$$\log(T_i) = -\psi_A R_i + \log(U_i). \quad (4)$$

Correspondingly, if $D_i = T_i$ in the treated arm and $D_i = 0$ in the placebo arm, then (1) gives

$$\log(T_i) = -\log(1 - \Delta) \times R_i + \log(U_i). \quad (5)$$

The proportional hazards model with the parameter ψ_H from (3) and the accelerated failure time model with the parameter ψ_A from (4) are alternatives for assessing the ITT effect. If U_i follows a Weibull distribution then model (4) is a Weibull regression model and $\kappa\psi_A = \psi_H$ where κ is the index parameter of the Weibull distribution.⁹ We write $\Delta_H = 1 - \exp(\psi_H)$ and $\Delta_A = 1 - \exp(\psi_A)$ for the structural parameters corresponding to (3) and (4). The equivalence of (4) and (5) under full compliance motivates Δ_A . The equivalence of (3) and (4) under the Weibull model motivates Δ_H . In a non-Weibull case Δ_H does not have direct interpretation in terms of Δ . Under full compliance the parameter Δ may be consistently estimated by $\hat{\Delta}_A$ either parametrically or semi-parametrically, but by $\hat{\Delta}_H$ only under the Weibull model. Under non-compliance all estimates will be attenuated relative to Δ .

5. THE 'AS-TREATED' (AT) ANALYSIS

A naive approach to estimating the incremental effect of treatment received is to use a time varying indicator $A_i(t)$ for being on active treatment at time t in the proportional hazards model

$$\lambda_i(t) = \lambda_0(t) \exp(\psi_{\text{AT}} A_i(t)). \quad (6)$$

Partial likelihood estimation of the risk ratio parameter $\exp(\psi_{\text{AT}})$ is straightforward.

We show in the Appendix that Δ can be estimated by the partial likelihood estimate $\hat{\Delta}_{\text{AT}} = 1 - \exp(\hat{\psi}_{\text{AT}})$ with a weakly consistent and asymptotically unbiased estimate $\hat{\psi}_{\text{AT}}$ from model (6) if (i) the treatment-free survival time U_i and the potential time on active treatment D_i^* (defined in the Appendix) vary independently of each other in the population and (ii) U_i has an exponential distribution. The rationale for the argument is that based on model (1) and assumption (i) the hazard at time t conditional on the event $I\{D_i < t\}$ is shown to have the general form

$$\lambda_{T|D}(t|d) = \begin{cases} \lambda_U(t - \Delta d) & \text{for those who have dropped out by time } t \\ (1 - \Delta)\lambda_U(t - \Delta t) & \text{for those still on active treatment at time } t \end{cases} \quad (7)$$

with $\lambda_U(t)$ the hazard function for treatment-free survival U . If assumption (ii) holds, that is, if $\lambda_U(t) = \lambda_u$ for all $t > 0$, then $1 - \Delta$ is retrieved from a proportional hazards model of the form (6) (see Appendix).

Expression (7) holds also when D_i^* is conditionally independent of U_i for a given linear combination of the covariates Z_i provided this linear combination is also included as an additional term in the hazard model (6). The linear combination in Z_i can be viewed as a propensity score.¹⁰ In general, however, we cannot fully expect to capture the dependence between potential treatment-free survival and time on active treatment by conditioning on Z_i , in which case $U_i \not\perp\!\!\!\perp D_i^* | Z_i$ and the AT approach will give biased results.

6. THE G-ESTIMATION (GE) ANALYSIS

Randomization guarantees that any variable measured at baseline will on average be balanced with respect to the treatment assignment. In particular, $U_i \perp\!\!\!\perp R_i$, that is, $\Pr(U_i \geq x | R_i = 0) = \Pr(U_i \geq x | R_i = 1)$ for all values of $x \geq 0$. Thus, a procedure for estimating Δ can be based on computing $U_i(\Delta)$ from (1) for given Δ , and then using a test for equality of the distribution of the treatment-free survival times between the two treatment arms. If the assumptions 1–4 in Section 3.3 are valid then $U_i(\Delta) \perp\!\!\!\perp R_i$ at the true value $\Delta = \Delta_0$. From a grid of values for Δ one chooses as estimate $\hat{\Delta}_{\text{GE}}$ the value for which the distribution of $U_i(\Delta)$ in the treated arm is in some sense closest to the observed distribution of $U_i(\Delta)$ in the placebo arm. One complication arises from the fact that due to censoring $U_i(\Delta)$ cannot always be computed from the observed data using model (1). In the next section we describe a new censoring variable $C_i(\Delta)$ and new observations $\{X_i(\Delta), \delta_i(\Delta)\}$ which can be computed, and for which $\{X_i(\Delta), \delta_i(\Delta)\} \perp\!\!\!\perp R_i$ at the true value $\Delta = \Delta_0$. Treating the pair $\{X_i(\Delta), \delta_i(\Delta)\}$ as the failure time and the censoring indicator we use the logrank test as a measure of equality for the distribution of the treatment-free survival in the two randomized arms. We write the logrank test statistic as $G(\Delta)$ which has an asymptotic standard normal distribution under $\Delta = \Delta_0$.⁶ In principle, a point estimate for Δ could be obtained by solving $G(\Delta) = 0$ for Δ , and approximate test-based $100(1 - \alpha)$ confidence intervals could be

Table I. Illustration of Δ -censoring in g-estimation ($C_i = 4$ in all cases)

Δ	Subject	Potential variables		Observed data				Quantities in Δ -censoring			
		U_i	D_i^*	R_i	D_i	X_i	δ_i	$U_i^x(\Delta)$	$C_i(\Delta)$	$X_i(\Delta)$	$\delta_i(\Delta)$
0.5	1	3	6	0	0	3	1	3	2	2	0
0.5	1	3	6	1	4	4	0	2	2	2	0
0.5	2	2.25	1.5	0	0	2.25	1	2.25	2	2	0
0.5	2	2.25	1.5	1	1.5	3	1	2.25	2	2	0
-1	3	6	2.5	0	0	4	0	4	4	4	0
-1	3	6	2.5	1	2.5	3.5	1	6	4	4	0
-1	4	3.75	0.25	0	0	3.75	1	3.75	4	3.75	1
-1	4	3.75	0.25	1	0.25	3.5	1	3.75	4	3.75	1

found as the range where $|G(\Delta)| \leq Z_{1-\frac{\alpha}{2}}$, where Z_p denotes the p th quantile of the normal distribution. Since $G(\Delta)$ is based on ranks, however, it is a step function in Δ and $G(\Delta) = 0$ cannot be solved exactly. Instead $\hat{\Delta}_{GE}$ is computed as the value where $G(\Delta)$ changes its sign and test-based confidence intervals are found accordingly.^{6, 11}

6.1. Δ -censoring

Censoring due to end of scheduled follow-up needs special care when using g-estimation. One cannot simply replace T_i by X_i in model (1) can calculate the respective value for U_i for a fixed Δ (denote this by $U_i^x(\Delta)$), because then for all $\Delta_0 \neq 0$ we have $U_i^x(\Delta_0) \not\perp R_i$ and assumption 3 of Section 3.3 is violated. Instead, we define a new censoring time $C_i(\Delta) = C_i$ if $\Delta \leq 0$ and $C_i(\Delta) = C_i(1 - \Delta)$ when $0 < \Delta < 1$. For given Δ we use $X_i(\Delta) = \min(U_i(\Delta), C_i(\Delta))$ and $\delta_i(\Delta) = I(X_i(\Delta) = U_i(\Delta))$ as the new follow-up time and censoring indicator. Both $X_i(\Delta)$ and $\delta_i(\Delta)$ can be calculated from the observed data and at $\Delta = \Delta_0$ we have $\{X_i(\Delta), \delta_i(\Delta)\} \perp R_i$ and the independence assumption 3 of Section 3.3 is restored. $C_i(\Delta)$ can be interpreted as the maximum $U_i(\Delta)$ that can be calculated which is smaller than C_i no matter how long individual i is on active treatment. This way of dealing with censoring was introduced by Robins and Tsiatis.⁶

Table I gives four illustrative examples of Δ -censoring in g-estimation. In all cases the censoring time C_i is fixed to 4. The first four rows illustrate what can happen when treatment is beneficial ($\Delta > 0$) and the last four rows describe situations when treatment is harmful ($\Delta < 0$). For illustration we show the potential quantities U_i and D_i^* (see Appendix) as well as the observed data in Table I. In the first case (row 1) we would observe the treatment-free survival time $X_i = U_i = 3$ if subject 1 was randomized to the placebo arm. In g-estimation his treatment-free survival time will be censored because of $C_i(0.5) = 2 < U_i$ and thus $X_i(0.5) = 2$. If the same person was randomized to the active treatment arm (row 2), he would be on active treatment until death at 6 which we cannot observe because the follow-up ends at 4. Instead we observe $D_i = X_i = C_i = 4$. He is a complier and in g-estimation his treatment-free time is censored at 2. Thus the distribution of $X_i(0.5)$ and $\delta_i(0.5)$ is independent of the treatment arm assignment. If we calculate $U_i^x(0.5)$ for these two cases we observe that $U_i^x(0.5) \perp R_i$. The next two rows (3 and 4) show that if treatment is beneficial and time on active treatment is short enough, then some extra censoring can happen also in the active treatment arm. Rows 5 and 6 show that when the treatment is harmful then some extra censoring can occur in the active treatment arm but not in

Table II. Illustration of g-estimation with a simple example ($C_i = 4$ in all cases)

Subject	U_i	Observed data				$\Delta = -0.5$		$\Delta = 0$		$\Delta = 0.5$	
		R_i	D_i	X_i	δ_i	$X_i(\Delta)$	$\delta_i(\Delta)$	$X_i(\Delta)$	$\delta_i(\Delta)$	$X_i(\Delta)$	$\delta_i(\Delta)$
1	5	1	4	4	0	4	0	4	0	2	0
2	4	1	4	4	0	4	0	4	0	2	0
3	3	1	2	4	1	4	0	4	1	2	0
4	2	1	1	2.5	1	3	1	2.5	1	2	1
5	1	1	0	1	1	1	1	1	1	1	1
6	5	0	0	4	0	4	0	4	0	2	0
7	4	0	0	4	1	4	1	4	1	2	0
8	3	0	0	3	1	3	1	3	1	2	0
9	2	0	0	2	1	2	1	2	1	2	1
10	1	0	0	1	1	1	1	1	1	1	1
						$G(-0.5)=1.065$	$G(0)=0.369$	$G(0.5)=0$			

the placebo arm. In this case $U_i^x(-1) \not\ll R_i$ but $X_i(-1) \ll R_i$. Whether treatment-free survival time is censored depends on the time spent on active treatment. Contrasting rows 6 and 8 we see that if time on active treatment is short enough then the treatment-free survival time will not be censored.

The general rule is that the survival times are censored on the U-scale in g-estimation when

$$\Delta D_i < T_i - (1 - \Delta)C_i \quad \text{if } 0 < \Delta < 1$$

$$\Delta D_i < T_i - C_i \quad \text{if } \Delta < 0.$$

If $\Delta = 0$ then the censoring status remains the same on the U-scale in g-estimation.

6.2. Illustrative example of GE-analysis

We illustrate the g-estimation procedure using a simple example. Table II gives hypothetical data for 10 subjects, five randomized to active treatment and five to placebo, respectively. The second column of Table II shows that the distribution of treatment-free survival time U_i is balanced with respect to treatment assignment. The observed data are given in columns 3 to 6. Subjects in the placebo arm do not receive the active treatment ($D_i = 0$). Subjects in the active treatment arm remain on active treatment the longer their treatment-free survival time. In this example the censoring time is fixed to be 4 time units for each subject. Table II reports values of $X_i(\Delta)$ and $\delta_i(\Delta)$ for three different values of Δ and the value of $G(\Delta)$. We see that some additional censoring is introduced in the active treatment arm when $\Delta = -0.5$ and in both treatment arms when $\Delta = 0.5$. In g-estimation we find $\hat{\Lambda}_{GE} = 0.5$ because $G(\Delta = 0.5) = 0$.

7. SIMULATION STUDY

We compare the performance of the ITT, AT and GE approaches under different settings for non-compliance with emphasis on the case when there are unmeasured confounders at baseline affecting both treatment-free survival time and time on active treatment. This extends the simulations by Mark and Robins.⁵

7.1. Simulation design

In the simulations we let the true treatment effect be either very harmful ($\Delta = -1$), ineffective ($\Delta = 0$) or very beneficial ($\Delta = 0.5$). Censoring time is fixed to 6 and the overall death rate in the placebo group is about 25 per cent. The overall drop-out rate in the active treatment group is set to either 0 per cent, 35 per cent or 50 per cent. We induce dependence by letting the treatment-free survival time and the potential time on active treatment depend on the same unmeasured confounder. In the simulation scheme described in detail below the overall death rate is controlled by the parameter β_0 , the overall drop-out rate by the parameter θ_0 and the dependence between treatment-free survival and time on active treatment by the parameters β_1 and θ_1 . The parameter β_0 is set to $\log(0.05)$ which corresponds to 25 per cent death rate in the placebo arm. The parameter θ_0 is set to either -20 , $\log(0.07)$ or $\log(0.12)$ which correspond to 0 per cent, 35 per cent and 50 per cent drop-out rate in the active treatment arm, respectively. We set $\beta_1 = \theta_1 = 1$ to let treatment-free survival time depend on the time on active treatment (outcome dependent case) and $\beta_1 = \theta_1 = 0$ to have independence (independence case). For all 18 possible combinations of the parameters we simulate a set of 1000 observations using the following scheme and we repeat the simulations 500 times:

1. Fix $C_i = 6$ for all $i = 1, \dots, 1000$.
2. Draw $Z_{i1} \sim N(0, 1)$ and $R_i \sim \text{Bernoulli}(\frac{1}{2})$. Define $Z_i = (1, Z_{i1})^T$.
3. Draw U_i using a linear transformation model¹²

$$\log(U) = -Z^T \beta + \varepsilon \quad (8)$$

with $\beta = (\beta_0, \beta_1)^T$ and ε from the extreme value distribution with distribution function $P(\varepsilon \leq x) = 1 - e^{-e^x}$.

4. Draw D_i^* using model (8) with $\beta = (\theta_0, \theta_1)^T$.
5. If $R_i = 1$ set $D_i = \min\left(D_i^*, \frac{U_i}{1 - \Delta}\right)$ and if $R_i = 0$ set $D_i = 0$.
6. Calculate T_i from the structural model (1).
7. Set $X_i = \min(T_i, C_i)$, $\delta_i = I(X_i = T_i)$ and $D_i = \min(D_i, X_i)$.

Note that the draws from the linear transformation model (8) are here exponentially distributed with rate parameter $\exp(Z^T \beta)$. The covariate Z_{i1} operates as a predictor for both U_i and D_i^* , thus causing dependence. Information on Z_{i1} is not used in any analysis.

7.2. Computation

The ITT and AT analyses are readily done using any survival analysis software. G-estimation involves finding the value of Δ for which the logrank test statistic is zero based on the data $\{X_i(\Delta), \delta_i(\Delta), R_i\}$ as defined in Section 6.1. A root finding procedure may be used such as the uniroot function in S-plus.¹³ All the calculations are done using S-plus version 3.4 Release 1 for a Sun SPARC (SunOS 5.3) computer.

7.3. Results

From each of the 18 simulations, we report in Table III the distribution for the estimated parameters over the 500 repetitions, their mean squared error (MSE), the median length of the 95 per cent confidence interval (CI), the proportion of hits in the 95 per cent CI, that is, how many

Table III. Results of the simulation study for ITT, AT and GE methods

Non-compliance setting	ITT				AT				GE			
	MSE	CI	Hits	Power	MSE	CI	Hits	Power	MSE	CI	Hits	Power
<i>Independence case</i>												
$\Delta = 0.5, \theta_0 = -20$	0.006	0.298	95.0	100.0	0.006	0.298	95.0	100.0	0.008	0.333	94.8	100.0
$\Delta = 0.5, \theta_0 = \log(0.07)$	0.014	0.331	80.2	97.8	0.006	0.313	94.0	99.8	0.010	0.408	95.6	98.0
$\Delta = 0.5, \theta_0 = \log(0.12)$	0.027	0.352	58.6	92.6	0.006	0.326	96.4	99.6	0.014	0.471	95.0	92.8
$\Delta = 0, \theta_0 = -20$	0.016	0.497	95.6	—	0.016	0.497	95.6	—	0.017	0.532	95.6	—
$\Delta = 0, \theta_0 = \log(0.07)$	0.015	0.499	96.0	—	0.016	0.509	96.0	—	0.023	0.644	96.0	—
$\Delta = 0, \theta_0 = \log(0.12)$	0.017	0.495	95.0	—	0.018	0.525	96.0	—	0.033	0.704	95.2	—
$\Delta = -1, \theta_0 = -20$	0.047	0.883	95.6	100.0	0.047	0.883	95.6	100.0	0.065	1.027	96.6	100.0
$\Delta = -1, \theta_0 = \log(0.07)$	0.065	0.815	86.8	100.0	0.047	0.885	95.8	100.0	0.073	1.138	95.0	100.0
$\Delta = -1, \theta_0 = \log(0.12)$	0.111	0.773	72.0	100.0	0.052	0.899	96.4	100.0	0.088	1.215	97.4	100.0
<i>Outcome dependent case</i>												
$\Delta = 0.5, \theta_0 = -20$	0.009	0.286	84.8	99.8	0.009	0.286	84.8	99.8	0.008	0.325	95.0	99.8
$\Delta = 0.5, \theta_0 = \log(0.07)$	0.041	0.331	27.8	90.0	0.015	0.225	60.0	100.0	0.014	0.485	94.6	90.4
$\Delta = 0.5, \theta_0 = \log(0.12)$	0.070	0.355	12.0	67.6	0.025	0.219	35.8	100.0	0.026	0.616	94.8	68.2
$\Delta = 0, \theta_0 = -20$	0.011	0.448	95.2	—	0.011	0.448	95.2	—	0.018	0.568	95.2	—
$\Delta = 0, \theta_0 = \log(0.07)$	0.014	0.450	94.4	—	0.062	0.358	37.0	—	0.041	0.775	94.6	—
$\Delta = 0, \theta_0 = \log(0.12)$	0.015	0.447	94.2	—	0.106	0.338	13.4	—	0.059	0.904	94.0	—
$\Delta = -1, \theta_0 = -20$	0.093	0.709	74.0	100.0	0.093	0.709	74.0	100.0	0.077	1.105	93.0	100.0
$\Delta = -1, \theta_0 = \log(0.07)$	0.212	0.646	35.2	98.6	0.362	0.578	8.4	90.4	0.109	1.305	95.4	98.6
$\Delta = -1, \theta_0 = \log(0.12)$	0.277	0.622	18.2	97.2	0.509	0.546	2.0	69.8	0.148	1.478	95.0	97.2

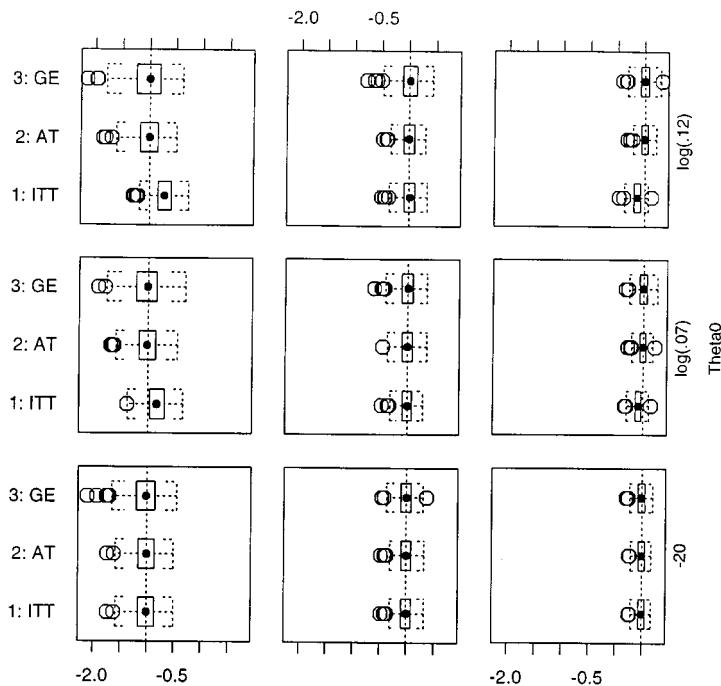


Figure 1. Box-whisker plots of the distribution of estimated Δ 's for the different methods compared under no outcome dependent non-compliance (that is, $\beta_1 = \theta_1 = 0$)

times in 100 cases the 95 per cent CI includes the true value of Δ , and the empirical power (Power) for rejecting the null hypothesis $\Delta = 0$. The power is reported only for those cases where the true treatment effect is not zero.

From Table III we see that the ITT approach performs well if the true treatment effect is null or if there is no drop-out. It gives narrower confidence intervals than AT or GE and also has proper coverage. The bias of ITT can immediately be seen if $\Delta_0 \neq 0$ and if there is drop-out (that is, $\theta_0 \neq -20$). The bias increases with the drop-out rate and is more pronounced with outcome dependent drop-out, which can be seen from the incorrect proportion of hits within the 95 per cent CI and from increased MSEs. If drop-out does not depend on the underlying treatment-free survival time, then the AT approach performs well for all values of the true treatment effect and the drop-out rate. AT has the correct coverage and MSE is low. With outcome dependent drop-out AT fails completely even if $\Delta_0 = 0$ and $\theta_0 \neq -20$. Note that by definition ITT is the same as AT if $\theta_0 = -20$. GE performs well in all situations. With outcome dependent drop-out GE has correct coverage in all 9 settings. It has wider confidence intervals throughout, which can be understood from the fact that no likelihood is specified and only the independence assumption between U_i and R_i is used. The power of the methods decreases when drop-out increases. Here the power is high because the treatment effects are large in magnitude and the overall death rate is quite high. Note, however, that the powers for ITT and GE are about the same, reflecting that even though the GE approach uses non-compliance information it does not increase the power against the null hypothesis when compared with the ITT approach. This reflects the fact that the

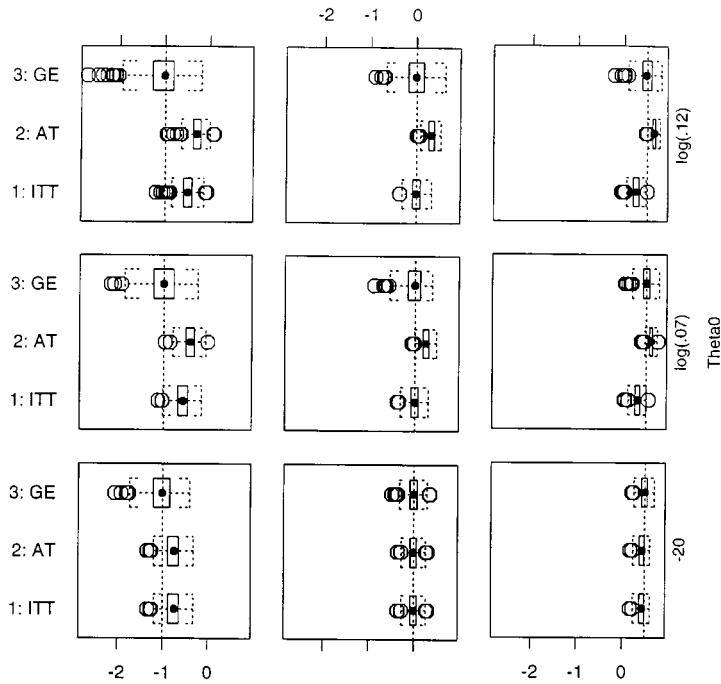


Figure 2. Box-whisker plots of the distribution of estimated Δ 's for the different methods compared under outcome dependent non-compliance (that is, $\beta_1 = \theta_1 = 1$)

GE test statistic and the ITT test statistics are algebraically equivalent at $\Delta = 0$. This robustness of the ITT approach is useful and it indicates that for testing purposes the ITT approach is valid even if it tends to underestimate the treatment effect.

Figures 1 and 2 show box-whisker plots of the distributions of the 500 estimated Δ 's under the different compliance settings. In Figure 1 we present the results from the independence case and Figure 2 shows the outcome dependent drop-out case. The lowest panel has no drop-out ($\theta_0 = -20$), the middle panel has drop-out around 35 per cent ($\theta_0 = \log(0.07)$) and the upper panel has a high drop-out rate ($\theta_0 = \log(0.12)$). The true treatment effect increases from left to right with values $\Delta_0 = -1, 0$ and 0.5 . Dashed lines indicate the true value of the treatment effect used for the respective panels. The distributions of the estimated parameters seem to be left-tailed which is due to the restriction $\Delta < 1$. Figure 1 shows that when drop-out is independent of outcome the empirical distribution of the estimated parameter is centred around the true value for AT and GE, but for ITT only in the case of no drop-out at all (lowest panel) or when there is no treatment effect. With outcome dependent drop-out (Figure 2) the attenuation of the estimated ITT effect is shown also when drop-out is absent, because treatment-free survival time and thus the observed survival time depends on an unmeasured confounder, which can be thought as a frailty term causing attenuation. The GE approach can cope with baseline unmeasured confounders because their distributions should be independent of the treatment assignment.

For completeness we repeated steps 1 to 7 of the simulation design in the non-Weibull case with $\Delta = -1, \beta_0 = \log(0.05), \theta_0 = \log(0.07)$ and either $\beta_1 = \theta_1 = 0$ or $\beta_1 = \theta_1 = 1$. We used the

Table IV. Drop-outs, deaths and person years accounted for active follow-up and total follow-up in the BC and NOBC groups of the ATBC Study

Group	<i>N</i>	Drop-outs	Deaths	Active follow-up*	Total follow-up*
NOBC	14573	3642	1718	72188	85018
BC	14560	3703	1853	71888	84714

* Person years

standard logistic distribution for ε in (8) instead of the extreme value distribution implying a log-logistic model for U and D instead of a Weibull model. For the independence case the proportions of hits within the 95 per cent CI for ITT, AT and GE were 68.6, 85.6, and 96.0, respectively. The bias of AT in the independence case can thus be seen if U is not from a Weibull distribution. For the outcome dependent drop-out the respective percentages were 34.6, 19.0 and 95.0.

8. BETA-CAROTENE AND TOTAL MORTALITY IN THE ATBC STUDY

We present the ITT, AT and GE analysis for data on all cause mortality from the ATBC Study. The design and the main results of the study have been described elsewhere.^{1,2} We only consider estimation of the magnitude of the effect of beta-carotene actually received on total mortality. We compare the beta-carotene group (BC) with the no beta-carotene (NOBC) group. For each participant we know the censoring time at the time of randomization. The median follow-up time was about 6 years. Table IV gives a summary of the number of drop-outs and deaths and person years for the two groups. Premature drop-out was defined as stopping clinic visits and being alive 120 days after the last attended visit. The drop-out rate was similar in the groups and out of 14,560 participants in the BC arm, 3703 withdrew from the study prematurely for reasons other than death. Compliance, estimated on the basis of residual capsule counts, was excellent while on the study.¹ Some 88 per cent of the participants took over 90 per cent of their prescribed capsules during active participation and only 4 per cent were poor compliers in this respect. The estimated overall capsule consumption was 93 per cent. Therefore, we have defined exposure to beta-carotene for simplicity as the number of days from randomization to the time of drop-out in the BC group. With this definition participants in the BC group received beta-carotene supplementation 84.9 per cent of their total time in the study (Table IV). The remaining 15.1 per cent is accounted for by the time after drop-out in the BC group. Thus, non-compliance is a relevant issue to address. With the above definition for drop-out D_i denotes the drop-out time which is set to zero in the NOBC group. D_i measures the total exposure to the beta-carotene supplementation during the study. For the AT approach $A_i(t) = 1$ if the participant in the BC group has not dropped out by time t and $A_i(t) = 0$ otherwise.

Baseline age, number of cigarettes smoked daily, years smoked, alcohol consumption, place of residence (urban/rural) and serum beta-carotene level were significant predictors for the duration of active treatment in the BC group (Table V). Baseline age, years smoked and serum beta-carotene level were also significant predictors for all cause mortality in the NOBC group (Table V). Thus there is no doubt that non-compliance is related to the endpoint through

Table V. Baseline predictors of mortality in the NOBC group and time on active treatment in the BC group

	NOBC*	BC†
	RR (Lower and upper 95% CI)‡	RR (Lower and upper 95% CI)‡
Age/10 years	2.46 (2.25, 2.69)	1.35 (1.27, 1.44)
Years smoked	1.06 (1.05, 1.07)	1.02 (1.09, 1.03)
No cigarettes	1.00 (0.99, 1.01)	1.01 (1.00, 1.02)
Place of residence	1.01 (0.91, 1.11)	1.13 (1.05, 1.21)
Alcohol§	1.04 (0.98, 1.09)	1.16 (1.13, 1.20)
Serum beta-carotene	0.83 (0.78, 0.89)	0.74 (0.70, 0.78)

* Mortality in the NOBC group

† Time on active treatment in the BC group

‡ Relative risk and confidence intervals estimated from Cox's proportional hazards model

§ Per one standard deviation (21.6 g/day)

|| Per one standard deviation (184 mmol/l)

Table VI. Estimated effects of beta-carotene supplementation on total mortality for the ITT, AT and GE approaches

Estimation approach	Estimate of Δ (lower and upper 95% CI)	Relative survival time* (lower and upper 95% CI)
ITT		
Cox's model (Δ_H)	-0.083 (-0.157, -0.014)	0.92 (0.86, 0.99)
Weibull model (Δ_A)†	-0.062 (-0.112, -0.011)	0.94 (0.89, 0.99)
Semi-parametric AFT (Δ_A)‡	-0.057 (-0.108, -0.010)	0.95 (0.90, 0.99)
AT (Δ_{AT})§	0.530 (0.494, 0.565)	2.13 (1.98, 2.30)
GE (Δ_{GE})	-0.079 (-0.154, -0.012)	0.93 (0.87, 0.99)

* Expresses relative survival time in the BC group compared to the NOBC group, that is $1/(1 - \Delta)$ † Index parameter estimated as $\hat{\kappa} = 1.28$

‡ Estimated with the GE approach

§ Relative risk adjusted for age, serum beta-carotene level and years smoked

measured baseline factors, but we do not exclude the possibility of other unmeasured baseline confounders, and thus do not expect the AT analysis to give valid estimates.

8.1. ITT results

Table VI summarizes the results of the total mortality analysis with the different approaches. The ITT analysis suggested a harmful effect of beta-carotene supplementation on total mortality. The relative risk estimate based on model (3) was 1.08 (95 per cent CI: 1.01 to 1.16), that is, 8 per cent higher death rate in the BC arm. In Table VI we present results both on the Δ -scale and as relative survival $(1 - \Delta)^{-1}$. The naive AT relative risk estimate when conditioning on baseline age, years smoked and serum beta-carotene level was 0.53. The apparent lower risk in the beta-carotene arm is an artefact due to the deaths being moved from the beta-carotene arm to the placebo arm at the time of drop-out. There is clearly residual dependence between drop-out and survival time even after conditioning on baseline factors, and the AT analysis is totally inappropriate here. We

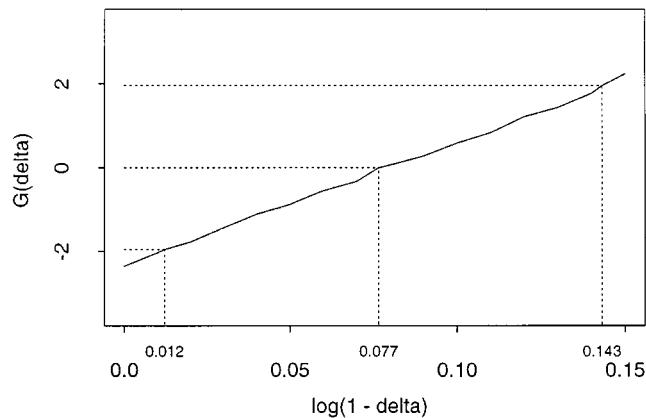


Figure 3. G-estimation analysis of total mortality in the ATBC Study. Results shown using scale $\log(1 - \Delta)$

analysed the ITT data also using an accelerated failure time model (4) which was defined by setting $D_i = \min(T_i, C_i)$ in the BC arm and $D_i = 0$ in the NOBC arm. The semi-parametric accelerated failure time ITT estimate (estimated using g-estimation) based on (4) suggest that survival time in the BC group is some 5.4 per cent shorter compared with the NOBC group. A parametric Weibull model yielded a similar effect of 5.9 per cent. The estimated index parameter was $\hat{\kappa} \approx 1.28$. We note that the ITT effects from a Cox model and from an AFT model are only comparable under a Weibull model. If the underlying true distribution for U_i , was Weibull, then the three estimated parameters would be approximately equal with $1/(1 - \hat{\Delta}_H) = 1/(1 - \hat{\Delta}_A)^{\hat{\kappa}}$. Using the Weibull $\hat{\Delta}_A$ we have $1/(1 - \hat{\Delta}_A)^{\hat{\kappa}} \approx 0.925$, which is close to the estimated effect from the Cox proportional hazards model (3).

8.2. GE results

The magnitude of the effect of beta-carotene actually received assessed by the GE analysis showed that the survival time would be some 7.4 per cent shorter for a subject always on beta-carotene supplementation as opposed to never being on beta-carotene supplementation (Table VI). In the BC group 136 observed deaths were censored due to g-estimation and none in the NOBC group. For illustration of the GE approach we present in Figure 3 values of $G(\Delta)$ plotted against $\log(1 - \Delta)$. The curve seems quite smooth which is due to the large number of deaths and participants in the study. With fewer events the curve would look more like a step function.

If the assumptions of Section 3.3 are valid then $U_i(\Delta_0)$ should be independent of any function involving R_i . Specifically, $U_i(\Delta_0) \perp\!\!\!\perp R_i | Z_i$ and thus $E(G(\Delta_0) | Z_i) = 0$ for all values of Z_i . An informal diagnostic check would be to assess whether $G(\hat{\Delta}_{GE}) = 0$ for different strata formed from the baseline variables, Z_i , when $\hat{\Delta}_{GE}$ is the estimated value of the overall structural parameter. Large absolute values of $G(\hat{\Delta}_{GE})$ within some strata would indicate that the structural model is incorrectly specified. A corresponding graphical procedure would be to plot the stratified Kaplan–Meier curves for both treatment arms for different strata of Z_i using data $\{X_i(\hat{\Delta}_{GE}), \delta_i(\hat{\Delta}_{GE})\}$. If the structural model is correct we would expect the curves to be identical at $\Delta = \Delta_0$ within each stratum. We performed a sensitivity analysis by stratifying on quartiles of baseline

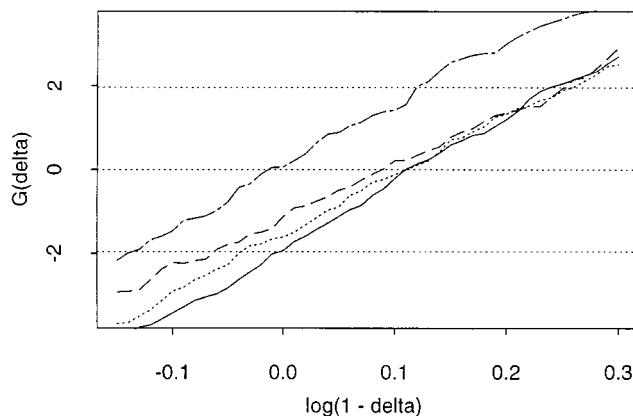


Figure 4. G-estimation analysis stratified by quartiles of baseline serum beta-carotene level (solid line = first, dotted = second, dashed = third, dashed with dots = fourth). Results shown using scale $\log(1 - \Delta)$

serum beta-carotene level. Values of $G(\Delta)$ evaluated at $\hat{\Delta}_{GE} = -0.079$ were -0.65 , -0.30 , -0.18 , and 1.26 , respectively, none of them being significant at the 5 per cent level. The Kaplan–Meier curves in each quartile overlapped (not shown) and thus did not indicate that the structural model is inadequate when checked from this angle. Figure 4 displays the respective $G(\Delta)$ curves. The three lowest quartiles give quantitatively similar results but the highest quartile seem to differ from the others. The estimated effects (95 per cent CI) using the relative survival time scale were 0.90 ($0.79, 1.00$), 0.89 ($0.78, 1.04$), 0.92 ($0.78, 1.07$) and 1.01 ($0.89, 1.15$), respectively. In the three lowest quartiles the treatment effect is harmful and of the same magnitude as the overall effect. Only in the highest quartile is treatment slightly beneficial but the wide confidence intervals do not exclude the possibility of a harmful effect.

Comparing the results of the GE approach with the accelerated failure time ITT effect we may conclude that the effect of treatment ‘as assigned’ is 5.4 per cent, and the effect of treatment ‘as received’ is 7.4 per cent, and there is clear attenuation of the ITT effect due to non-compliance, however, it does not change the overall conclusion of the study.

9. DISCUSSION

To estimate the magnitude of the effect of treatment actually received on a survival endpoint in randomized studies the possible mechanism causing non-compliance needs to be addressed. The ITT analysis often gives estimates biased towards the null and the AT approach can be seriously misleading when outcome dependent non-compliance is present. The ITT approach is valid for testing purposes and if a true treatment effect exists then provided the study has adequate power the intention-to-treat test would reject the null hypothesis. The GE approach or g-estimation introduced by Robins⁴ and Robins and Tsiatis⁶ offers an alternative that provides valid estimates under a different set of assumptions even when outcome dependent drop-out is present. GE closely relates to the instrumental variable methods which were recently applied in the context of non-compliance in randomized studies.³ Mark and Robins⁵ gave simulation results for the performance of GE as compared with the ordinary ITT logrank test in a randomized study. We

report an extended simulation study which explicitly shows that GE will work even under outcome dependent non-compliance when the underlying structural model is correct. One weakness of the GE approach is that extra censoring is introduced due to the estimation method and thus it induces loss of power. In the GE approach presented in this paper we have information on subjects until they die or are censored at some fixed time point regardless of their drop-out status. In many studies, however, information on deaths and other events cannot be obtained after loss to follow-up. This is the case for example if no reliable registers are available for endpoint ascertainment. In such cases the proposed GE approach cannot be directly applied and it needs to be modified to account for possible censoring by competing risks (that is, random loss to follow-up).¹⁴

Another weakness of GE is that it entirely relies on the structural model. Model (1) assumes that the treatment effect Δ is constant in time and constant with respect to any other covariate; this may be unrealistic. Time-by-treatment or treatment-by-covariate interactions as well as random noise may exist. We proposed informal methods for checking the adequacy of the structural model, which extend to censored survival data the method suggested by Goetghebeur and Lapp¹⁵ for continuous normal responses. We also use this theory on real data from the ATBC Study. However, a formal justification for the diagnostic tools presented here would be useful.

Robins and Tsiatis⁶ and Mark and Robins⁵ also indicate how the one parameter structural model may be extended to the multi-parameter case. Multi-parameter structural failure time models have been fitted by Robins and Greenland¹⁶ and White and Goetghebeur.¹⁷ There is also an application for linear models with a continuous normal outcome.¹⁵ White and Goetghebeur applied the GE approach in a more complex situation where two active treatments are compared in an elderly hypertension trial.¹⁷ Work is in progress for extending g-estimation for the ATBC Study data to include a treatment modifying baseline term $\Delta_2 D_i \times Z_i$ in the structural model (1).

APPENDIX

Let $U_i \sim f_U(u)$ denote the treatment-free survival time and $D_i^* \sim f_{D^*}(d)$ the potential time on the active treatment for subject i .

In the active treatment arm we make the following transformation from the potential event times (U_i, D_i^*) to the observed event times (T_i, D_i) :

$$T_i = \begin{cases} U_i + \Delta D_i^* & \text{for } D_i^* < \frac{U_i}{1 - \Delta} \\ \frac{U_i}{1 - \Delta} & \text{for } D_i^* \geq \frac{U_i}{1 - \Delta} \end{cases}$$

and

$$D_i = \begin{cases} D_i^* & \text{for } D_i^* < \frac{U_i}{1 - \Delta} \\ \frac{U_i}{1 - \Delta} & \text{for } D_i^* \geq \frac{U_i}{1 - \Delta} \end{cases}$$

with $-\infty < \Delta < 1$ because T_i must be positive. Here $D_i \leq T_i$ and the construction satisfies the model $T_i = U_i + \Delta D_i$ introduced in (1).

The joint density for observed T_i and D_i is given by

$$f_{T,D}(t,d) = \begin{cases} f_{U,D^*}(t - \Delta d, d) & \text{for } d < t \\ (1 - \Delta) \int_t^\infty f_{U,D^*}(t - \Delta t, d^*) dd^* & \text{for } d = t \end{cases}$$

where $1 - \Delta$ is the Jacobian of the transformation for those still on active treatment at death t . We denote by S_X and λ_X the survival and hazard functions of a random variable X . Under the assumption $U_i \perp\!\!\!\perp D_i^*$ the joint density simplifies to

$$f_{T,D}(t,d) = \begin{cases} f_U(t - \Delta d) f_{D^*}(d) & \text{for } d < t \\ f_U(t - \Delta t) S_{D^*}(t) (1 - \Delta) & \text{for } d = t \end{cases}$$

which is a proper density, that is, $\int_0^\infty \int_0^t f_{T,D}(t,d) dd dt = 1$. This gives the marginal density for D_i

$$\begin{aligned} f_D(d) &= \int_d^\infty f_{T,D}(t,d) dt \\ &= f_U(d - \Delta d) f_{D^*}(d) \left[\frac{(1 - \Delta) \lambda_U(d - \Delta d) + \lambda_{D^*}(d)}{\lambda_U(d - \Delta d) \lambda_{D^*}(d)} \right] \end{aligned}$$

and the survival distribution for D_i

$$S_D(t) = \int_t^\infty f_D(d) dd = \int_{t(1-\Delta)}^\infty \int_t^\infty f_{U,D^*}(u, d^*) dd^* du = S_U(t - \Delta t) S_{D^*}(t).$$

Now, the conditional distribution for T_i , for given $D_i = d$ is

$$f_{T|D}(t|d) = \begin{cases} \frac{f_U(t - \Delta d) \lambda_{D^*}(d)}{S_U(d - \Delta d) (1 - \Delta) \lambda_U(d - \Delta d) + \lambda_{D^*}(d)} & \text{for } d < t \\ \frac{(1 - \Delta) \lambda_U(t - \Delta t)}{(1 - \Delta) \lambda_U(t - \Delta t) + \lambda_{D^*}(t)} & \text{for } d = t \end{cases}$$

which is also a proper density that integrates to 1.

Finally, the hazard for T_i at time t conditional on the event {drop-out has happened before time t } = $I\{D_i < t\}$ equals

$$\lambda_{T|D}(t|d) = \frac{f_{T|D}(t|d)}{\int_t^\infty f_{T|D}(t|d) dt} = \lambda_U(t - \Delta d) \quad (9)$$

for those who have dropped out before time t , that is, $D_i < t$ and

$$\lambda_{T|D}(t|d) = \frac{f_{T,D}(t,t)}{\int_t^\infty f_D(t) dd} = (1 - \Delta) \lambda_U(t - \Delta t) \quad (10)$$

for those who are still on active treatment at time t , that is, $D_i \geq t$.

In the placebo arm we observe U_i if there was no censoring and $D_i = 0$ by definition. The hazard at time t is $\lambda_U(t)$ corresponding to the conditional hazard (9) with $D_i = 0$.

Note, that if $U_i \sim \exp(\lambda_u)$, then (9) and (10) reduce to

$$\lambda_{T|D}(t|d) = \begin{cases} \lambda_u & \text{for those who dropped out before time } t \\ (1 - \Delta) \lambda_u & \text{for those still on active treatment at time } t \end{cases}$$

and $1 - \Delta$ corresponds to the relative risk parameter in the proportional hazards model with a time dependent indicator, $A_i(t) = I(D_i \geq t)$, as an explanatory variable.

ACKNOWLEDGEMENTS

This work was supported by the grants numbers 31116 and 33495 from the Academy of Finland, grants numbers 3498 and 3816 from the Yrjö Jahnsson Foundation, and by NIH grant number GM-29745 and by contract number NOI-CN-45165 with the National Cancer Institute. We thank the referees for their helpful comments and Marshall Joffe, Louise Ryan and Shu Zhang for the informal Spring 1996 seminars on g-estimation at the Department of Biostatistics at Harvard University. The ATBC Study Group at the National Public Health Institute in Finland is acknowledged for initiating this research and for letting us use their data.

REFERENCES

1. The Alpha-Tocopherol Beta Carotene Cancer Prevention Study Group. 'The Alpha-Tocopherol, Beta Carotene lung cancer prevention study: design, methods, participant characteristics, and compliance', *Annals of Epidemiology*, **4**, 1–9 (1994).
2. The Alpha-Tocopherol Beta Carotene Cancer Prevention Study Group. 'The effect of vitamin e and beta carotene on the incidence of lung cancer and other cancers in male smokers', *New England Journal of Medicine*, **330**, 1020–1035 (1994).
3. Angrist, J. D., Imbens, G. W. and Rubin, D. B. 'Identification of causal effects using instrumental variables (with comments)', *Journal of the American Statistical Association*, **91**, 444–472 (1996).
4. Robins, J. M. 'The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies', in *Health Service Research Methodology: A Focus on AIDS*, NCHSR, U.S. Public Health Service, 1989, pp. 113–159.
5. Mark, S.D. and Robins, J. M. 'A method for the analysis of randomized trials with compliance information: an application to the multiple risk factor intervention trial', *Controlled Clinical Trials*, **14**, 79–97 (1993).
6. Robins, J. M. and Tsiatis, A. A. 'Correcting for non-compliance in randomized trials using rank preserving structural failure time models', *Communications in Statistics*, **A**, **20**(8), 2609–2631 (1991).
7. Cox, D. R. 'Regression models and life tables (with discussion)', *Journal of the Royal Statistical Society, Series B*, **34**, 187–220 (1972).
8. Andersen, P. K. and Gill, R. D. 'Cox's regression model for counting processes: a large sample study', *Annals of Statistics*, **10**, 1100–1120 (1982).
9. Cox, D. R. and Oakes, D. *Analysis of Survival Data*, Chapman and Hall, London, 1984.
10. Rosenbaum, P. R. and Rubin, D. B. 'The central role of the propensity score in observational studies', *Biometrika*, **70**, 41–55 (1983).
11. Tsiatis, A. A. 'Estimating regression parameters using linear rank test for censored data', *Annals of Statistics*, **18**, 354–372 (1991).
12. Cheng, S. C., Wei, L. J. and Ying, Z. 'Analysis of transformation models with censored data', *Biometrika*, **82**, 835–845 (1995).
13. MathSoft Data Analysis Product Division. *S-plus Version 3.4 for Unix Supplement*, MathSoft, Seattle, 1996.
14. Robins, J. M. 'Analytic method for estimating hiv treatment and cofactor effects', in *Methodological Issues in AIDS Mental Health Research*, Plenum Press, New York, 1993, pp. 213–290.
15. Goetghebeur, E. and Lapp, K. 'The effect of treatment compliance in a placebo-controlled trial: regression with unpaired data', *Applied Statistics*, **46**, 351–364 (1997).
16. Robins, J. M. and Greenland, S. 'Adjusting for differential rates of prophylaxis therapy for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial', *Journal of the American Statistical Association*, **89**, 737–749 (1994).
17. White, I. R. and Goetghebeur, E. J. T. 'Clinical trials comparing two treatment policies: which aspects of the treatment policies make a difference', *Statistics in Medicine*, **17**, 319–339 (1998).