Accelerated failure time models for non-ignorable non-compliance in randomized studies

Pasi A. Korhonen

Department of Nutrition
National Public Health Institute
and
Division of Biometry
Rolf Nevanlinna Institute

Faculty of Science
University of Helsinki

Academic Dissertation for the Degree of Doctor of Philosophy

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in Auditorium XII, University Main Building, on June 21th, 2000, at 12 o'clock noon.

Accelerated failure time models for non-ignorable non-compliance in randomized studies

Pasi A. Korhonen

Department of Nutrition
National Public Health Institute
and
Division of Biometry
Rolf Nevanlinna Institute

Research Reports A31
June 2000

Rolf Nevanlinna Institute
P.O. Box 4 (Yliopistonkatu 5)
FIN-00014 University of Helsinki, Finland

Printed version
ISBN 952-9528-57-4
ISSN 0787-8338
YLIOPISTOPAINO
HELSINKI 2000

PDF version
ISBN 951-45-9424-X
http://ethesis.helsinki.fi/
HELSINGIN YLIOPISTO
HELSINKI 2000

Acknowledgements

I want to express my sincere gratitude to my supervisor, Professor Juni Palmgren for her guidance and fruitful collaboration during the past five years. The initiation of this work would not have been possible without her incredible intuition to see and motivate relevant research problems. Her international connections gave me an opportunity to start my PhD studies at the Harvard School of Public Health in Boston during the academic year 1995-96.

The year at Harvard was an important building block for me. I first learned the fundamentals of non-compliance in an informal seminar at Harvard. I want to thank Professor Louise Ryan, Professor Marshall Joffe, Dr Shu Zhang and Juni, the participants of this seminar, for several hours of presentations and inspiring discussions which helped me to sort out the statistical concepts. The ground work of the first paper of this thesis was done during that year. I am very grateful to Professor Nan Laird not only for hosting the academic visit but also for her comments and encouragement during all stages of writing the first paper.

Professor Els Goetghebeur and Tom Loeys from the University of Ghent deserve my honest thanks. I have enjoyed working with them and find the intensive collaboration with them most productive. I want to give special thanks to Tom and his family for their hospitality during my visit in Ghent.

Professor Elja Arjas and Professor Jukka Sarvas gave me an opportunity to work at the Division of Biometry of the Rolf Nevanlinna Institute. The people at the Rolf Nevanlinna Institute created an unforgettable atmosphere for doing research and for enjoying doing it.

This project started and also ended at the National Public Health Institute. I am proud to be part of the Setti group. The Setti Study or in other words the ATBC Study was the original inspiration of the current research topic. The Setti people have created an enjoyable and functioning working environment. Especially I want to thank Dr Jarmo Virtamo for having understanding and for giving an opportunity and time to develop statistical methods for real life problems.

Professors Scott Zeger and Timo Hakulinen acted as pre-examiners of my thesis and I want to thank them for providing helpful comments and suggestions which helped to improve the text.

I wish to thank the Academy of Finland, the Yrjö Jahnsson Foundation and the Finnish Cancer Organisations for their financial support which made this work possible. The journals *Statistics in Medicine* and *Lifetime Data Analysis* are acknowledged for their permissions to reprint the original papers in this thesis.

Finally, I wish to give my warmest and loving thanks to my wife Tuula and to my children Aku, Iisa and Aleksi. Their encouragement and full support at all times helped me to carry out this project. They are the light of my life and I dedicate this work to them.

Pasi Korhonen -- Espoo, May 2000

List of original publications

This thesis is based on the following original articles which are referred to in the text by their Roman numerals:

I Korhonen P., Laird N. M., Palmgren J. (1999) Correcting for non-compliance in randomized trials: An application to the ATBC Study, *Statistics in Medicine*, **18**, 2879-2897, 1999

II Korhonen P., Palmgren J. (2000) Effect modification in a randomized trial under nonignorable noncompliance: An application to the ATBC Study. To appear in *Applied Statistics*, 30 pages

III Korhonen P., Loeys T., Goetghebeur E., Palmgren J. (2000) Vitamin A and infant mortality: Beyond intention-to-treat in a randomized trial, *Lifetime Data Analysis*, **6**, 107-121, 2000

IV Korhonen P. (2000) Joint estimation of effects of beta-carotene supplementation and smoking cessation on overall survival and on time to lung cancer diagnosis in the ATBC Study. Submitted for publication, 25 pages

Contents

1 Introduction **9**

2 The Rubin causal model and extensions **10**

2.1 Binary outcome and binary or categorical compliance 11

2.2 Continuous outcome and continuous or binary compliance 12

2.3 Survival outcome and continuous compliance 12

3 Estimation of the parameters of the accelerated failure time model **14**

3.1 The estimating procedure 14

3.2 Censoring by end of follow-up 15

3.3 Censoring by competing risks 15

3.4 Modification for missing covariates in the accelerated failure time model 16

3.5 Previous applications of the accelerated failure time models 17

4 Summaries of the original papers **17**

4.1 Paper I. Correcting for non-compliance in randomized trials: An application to the ATBC Study 17

4.2 Paper II. Effect modification in a randomized trial under nonignorable noncompliance: An application to the ATBC Study 19

4.3 Paper III. Vitamin A and infant mortality: Beyond intention-to-treat in a randomized trial 20

4.4 Paper IV. Joint estimation of effects of beta-carotene supplementation and smoking cessation on overall survival and on time to lung cancer diagnosis in the ATBC Study 20

5 Concluding remarks **21**

References **22**

1 Introduction

We consider a randomized double-blind placebo-controlled study where the subjects are randomized to an active treatment regime ($R_i = 1$) or to placebo ($R_i = 0$). The aim of the study is to test whether the active treatment can reduce the risk of a failure event such as cancer or death when compared to placebo. At enrollment a set of baseline covariates (Z_i) are observed for each subject i and in general the treatment assignment R_i can depend on Z_i. The outcome of interest (T_i) is defined as the time from randomization to the event of interest, say death. The subjects are monitored for their outcome until a pre-specified end of follow-up (C_i) and for each subject $X_i = \min(T_i, C_i)$ and $\Delta_i = I(X_i = T_i)$ are observed. During follow-up the subjects may visit a study center at regular intervals and a number of time-dependent covariates ($L_i(t)$) are recorded. Treatment regime is specified in the study protocol. It may be defined, for instance, to take one capsule of the assigned treatment once a day until the end of follow-up C_i. Some subjects fail to adhere to their intended treatment regime and stop taking the assigned capsule at some point after enrollment. A subject's actual receipt of active treatment is recorded by $A_i(t)$ which equals 1 if the subject took the active treatment at time t and 0 otherwise. We assume that subjects in the placebo arm do not have access to the active treatment, that is in the placebo arm $A_i(t) = 0$ for all i and $t > 0$. We call subject i in the active treatment arm a complier if $A_i(t) = 1$ for all $t \leq X_i$ and a non-complier otherwise.

The gold standard for the analysis of a randomized study involving humans is to use the intention-to-treat principle where the event rates in the two randomized groups are compared regardless of subsequent adherence to the assigned treatment regime. This provides a valid test for the null hypothesis of no treatment effect because randomization implies that the two groups are comparable on average with respect to their observed and unobserved baseline characteristics that may influence the outcome. Thus under the null hypothesis any differences between the observed event rates in the two treatment groups are due to chance via the random assignment of the treatment. Under the alternative, the difference between the observed event rates is an unbiased estimate of the effect of the treatment assignment on the outcome. This effect is generalizable to situation where compliance and factors influencing compliance are similar as in the study settings. The intention-to-treat estimate thus quantifies the effect of the treatment policy under incomplete compliance. The intention-to-treat estimate is often referred to as the estimate of the *effectiveness* of the treatment policy [1]. An analysis by the intention-to-treat principle should be the primary analysis of any randomized trial. It fulfils the *pragmatic aim of the trial* [2] of estimating the treatment effect when faced with incomplete compliance. The trial conditions may not be met in reality for several reasons. For instance the knowledge of an equal chance to receive placebo may result in poorer compliance in the trial settings, or participation in a trial may enhance compliance within the trial. On the other hand, the published trial results may yield better compliance when the treatment is used in general practice. Hence the estimated treatment effect may not be reproduced outside the trial settings.

When the compliance information $A_i(t)$ is available it is of interest to investigate the effect of the treatment actually received on the outcome. This effect, which represents the biological impact of the treatment regime, is referred to as the *efficacy* of the treatment [1]. The *explanatory aim of the study* [2] could be to quantify the magnitude of the treatment effect if everyone in the study adhered to their assigned treatment regime. This *biological* or *causal effect* of the treatment is more likely to be reproducable outside the trial settings. The treatment actually received, $A_i(t)$, is a post-randomization variable and may depend on the health status of the subject through measured and unmeasured covariates. Thus the compliers and non-compliers are likely to be inherently different with respect to their outcome inducing non-ignorable non-compliance [3]. Therefore, an attempt to estimate the efficacy using for example a Cox model $\lambda_i(t \mid A_i(t)) = \lambda_0(t)\exp(\beta \times A_i(t))$ with time-dependent covariates fails because those receiving active treatment at time t are prognostically different from those not receiving active treatment at time t.

This thesis consist of four original research papers which are referred to as Paper I - Paper IV. It builds on the work by Robins and co-authors [4, 5, 6, 7, 8] who use accelerated failure time models [9] to quantify the treatment action. Each paper focusses on a particular application. In Papers I and II we estimate the effect of beta-carotene supplementation actually received and its interactions with baseline covariates on survival in the Alpha-Tocopherol Beta-Carotene (ATBC) Cancer Prevention Study [10, 11]. In Paper III the focus is on the estimation of the effect of one dose of vitamin A on subsequent 4 month mortality in children under 6 months of age in a randomized, double-blind, placebo-controlled community trial in Nepal [12]. In Paper IV we estimate jointly the effects of beta-carotene supplementation, smoking cessation and their interaction on overall survival and on time to lung cancer diagnosis in the ATBC Study. Paper IV does not directly adress the issue of non-compliance, but smoking cessation as a post-randomization covariate, has a similar role as compliance when its effect on a survival outcome is explored. Thus similar models and estimation principles as in Papers I-III are used in Paper IV to provide an approach for addressing a different subject matter question.

2 The Rubin causal model and extensions

Statistical methods that account for non-ignorable non-compliance or other post-randomization covariates in a randomized trial build on the Rubin causal model as defined by Rubin [13, 14] and Holland [15]. As a key concept a subject may have several potential outcomes depending on the treatment assignment and on the receipt of treatment. An outline is given below for the statistical theory starting from binary outcomes with binary non-compliance and extending to more complicated situations with continuous and survival outcomes and continuous non-compliance. In all approaches the estimation relies on the distributional assumptions implied by randomization. The more complex models need additional assumptions to make the parameters of interest estimable from the observed data.

2.1 Binary outcome and binary or categorical compliance

Tarwotjo et al. [16] first considered the issue of estimating the effect of treatment received in an analysis of a randomized vitamin A trial with a control group where both the response and non-compliance were binary. In Tarwotjo et al. [16] the response Δ_i is death during a fixed follow-up period. Binary non-compliance arises from the study design in which a single oral dose of vitamin A is randomly assigned to children in a number of randomly selected villages, while children in the remaining villages serve as controls. A child assigned to receive the vitamin A dose either actually reveices or does not receive it, that is either $A_i = A_i(t) = 1$ or $A_i = A_i(t) = 0$ for all t. Sommer and Zeger [17] give a more detailed mathematical account of this approach. Randomization implies that the expected compliance rates in the two treatment arms are the same and thus they can construct an inferred group of compliers in the control arm who would have complied if assigned to active treatment. Assuming that the non-compliant subgroups in the two arms have the same expected mortality rate they derive an estimator for the efficacy which avoids the inherent selection bias that would arise if only the compliant subgroup in the treatment arm was compared with the control arm. Cuzick el al. [18] extend the Sommer and Zeger approach to allow contamination within the control group, that is some subjects in the control arm may receive the active treatment. Assuming that the efficacy is the same in the contaminated subgroup as in the compliant subgroup they derive an estimator for the efficacy using a standard binomial likelihood for the observed data. They also discuss how to use their approach with censored survival data.

Goetghebeur and Molenberghs [19] extend the Sommer and Zeger approach for ordered categorical compliance. They make a monotone dose-response assumption and assume that the zero dose outcome in the treatment group equals the outcome in the placebo group and they call this "the placebo link assumption". If there are more than two compliance categories the multinomial parameters of the joint distribution of non-compliance and response are not estimable without further restrictions on the parameters. Goetghebeur and Molenberghs use a class of models where they restrict all global odds ratios to coincide to make the model parameters estimable.

Angrist et al. [20] use the instrumental variable approach together with potential variables to estimate efficacy. They make five explicit assumptions: (i) there is no interference between subjects which is called the stable unit treatment value assumption in the Rubin causal model [15], (ii) the treatment is assigned at random, (iii) the exclusion restriction holds which requires that the treatment assignment has an effect on the outcome only through its effect on the treatment received, (iv) the treatment assignment has a nonzero effect on the receipt of treatment, and (v) the monotonicity assumption holds which requires that $A_i(R_i = 1) \geq A_i(R_i = 0)$ where $A_i(R_i = r)$ denotes the treatment reveived if randomized to group $r = 0$ or 1. Under these assumptions R_i is called an instrumental variable which is used to derive an estimator of the efficacy parameter which is interpreted as the average causal effect of the treatment reveived in the subgroup of compliers. Note that assumptions (i) - (v) are implicit also in the approaches above.

Imbens and Rubin [21] present Bayesian inferential methods for efficacy estimands from placebo-

controlled trials with binary non-compliance under assumptions similar to those as in Angrist et al. [20], but with a weaker version of the exclusion restriction. They assume that the treatment assignment is unrelated to the potential outcomes for those who would either always take or never take the active treatment.

2.2 Continuous outcome and continuous or binary compliance

An early attempt to consider efficacy in the context of randomized trials with a continuous compliance measure and continuous response was presented by Efron and Feldman [22]. They focussed on estimation of the average gain of an active treatment over placebo at a given exposure level using a fully parametric model. They made the restricting assumption that compliance is an inherent attribute of the subject which is determined prior to randomization and they required that adherence to the assigned treatment in the control arm and in the active treatment arm are similar. Several arguments against this assumption were raised in the discussion of the paper [22]. Albert and DeMets [23] show in simulations that even moderate deviation from the comparability of the adherence to the assigned treatment in the two arms can introduce severe bias. The Efron and Feldman approach further suffers from an unidentifiability problem which they can partly reduce by deriving a lower bound for the efficacy estimate at full compliance [22]. Goetghebeur and Shapiro [24] generalize the Efron and Feldman approach by allowing the expected placebo compliance to have shifted from the treatment compliance and they let the shift depend on measured side-effects while on active treatment and possibly also on baseline covariates. By this extra modelling assumption the unidentifiability problem of the Efron and Feldman approach is resolved. Goetghebeur and Lapp [25] present an application of the parametric nested mean models introduced by Robins [26] and Robins and Greenland [27] to estimate the effect of treatment at different levels of compliance. Based on the randomization assumption they obtain closed form estimates for the parameters of the structural model and their covariance matrix. Very recently Frangakis and Rubin [28] addressed the problem of estimating the efficacy in the presence of binary non-compliance and subsequent missing outcomes.

2.3 Survival outcome and continuous compliance

Robins and Tsiatis [5] first presented a semi-parametric counter-factual approach to account for non-compliance in randomized trials with a survival outcome. Following Rubin [14] and Holland [15] each subject in a placebo-controlled randomized trial has at least the following four potential survival times conceptually defined prior to randomization: $T_i\{R_i = 1\}$, $T_i\{R_i = 0\}$, $T_i\{R_i = 1, A_i(t) = 1 \text{ for all } t > 0\}$ and $T_i\{R_i = 0, A_i(t) = 0 \text{ for all } t > 0\}$ where $T_i\{R_i = r\}$ denotes the survival time if subject i would be randomized to treatment arm $r = 0, 1$ irrespective of compliance, $T_i\{R_i = 1, A_i(t) = 1 \text{ for all } t > 0\}$ denotes the survival time if subject i would be randomized to active treatment and he would always receive active treatment and $T_i\{R_i = 0, A_i(t) = 0 \text{ for all } t > 0\}$ denotes the survival time if subject i would be randomized to reveice placebo and he would

never receive active treatment. An effectiveness measure of treatment assignment for subject i may be defined as the difference $T_i\{R_i = 1\} - T_i\{R_i = 0\}$. However, the "fundamental problem" in a randomized trial is that for subject i we can never jointly observe all potential survival times of interest. The "statistical solution" for the effectiveness estimate is to use the average causal effect of treatment assigment $E(T_i\{R_i = 1\}) - E(T_i\{R_i = 0\})$ where the expectation is taken over the whole population. Randomization implies that these expectations are estimable from the two treatment arms and thus gives the intention-to-treat estimator. When one attempts to capture the biological efficacy of active treatment one would like to compare the outcomes under ideal conditions, namely by comparing $T_i\{R_i = 1, A_i(t) = 1 \text{ for all } t > 0\}$ with $T_i\{R_i = 0, A_i(t) = 0 \text{ for all } t > 0\}$. Because $A_i(t)$ as a post-randomization covariate cannot be assigned or controlled by study design, we can only observe one survival time $T_i = T_i\{R_i = r_i, A_i(t) = a_i\}$ where r_i is the assigned treatment and a_i is the particular realization of the actual receipt of treatment. Modelling assumptions are needed to recover the efficacy parameter from the observed data. Robins and Tsiatis [5] defined the potential outcome $U_i = T_i\{R_i, A_i(t) = 0 \text{ for all } t > 0\}$ as the survival time from enrollment to the study if active treatment had been witheld throughout the study. This "treatment-free survival time" or "placebo prognosis" is linked to the observed survival time and to the observed receipt of treatment through a parametric accelerated failure time model

$$U_i \overset{d|Z_i}{=} \int_0^{T_i} \exp\{\psi A_i(t)\} ds, \tag{1}$$

where the parameter ψ quantifies the efficacy of the treatment and $\overset{d|Z_i}{=}$ means equality in distribution conditional on Z_i. Model (1) describes how U_i is related to the observed T_i if subject i receives active treatment according to the observed process $A_i(t)$. The factor e^{ψ} can be interpreted as the relative increase or decrease in survival if never on active treatment compared to if always on active treatment. When the active treatment is not available outside the trial settings then we have $T_i\{R_i = 0\} = U_i$ and thus we would observe U_i in the placebo arm if there were no censoring. In the active treatment arm we only observe U_i for uncensored subjects who do not receive active treatment at all. In general the treatment-free survival time U_i may depend on baseline covariates Z_i and $A_i(t)$ may depend on U_i either directly or through measured and unmeasured covariates. Non-ignorable non-compliance in the sense of Little and Rubin [3] is thus introduced. Model (1) can be extended to allow the treatment effect to depend on baseline covariates

$$U_i \overset{d|Z_i}{=} \int_0^{T_i} \exp\{\psi_1 A_i(t) + \psi_2 Z_{ij} \times A_i(t)\} ds \tag{2}$$

where $\psi = (\psi_1, \psi_2)^T$ quantifies the treatment effect for a given value of the jth baseline covariate Z_{ij}. Randomization implies that the distribution of U_i is conditionally independent of the treatment assigment given the baseline covariates, or equivalently

$$U_i \perp\!\!\!\perp R_i \mid Z_i. \tag{3}$$

This key condition is used as a basis for estimation.

3 Estimation of the parameters of the accelerated failure time model

3.1 The estimating procedure

We briefly describe the main ideas of the estimation procedure developed in this thesis. Let us first assume that there is no censoring. Three assumptions are needed for estimation of ψ in models (1) and (2): (i) the structural model (1) or (2) correctly captures the treatment action, (ii) physical randomization takes place at baseline and the assignment mechanism is known, and (iii) U_i is unaffected by the survival experiences of other individuals. The last assumption is relaxed in Paper III where the survival times of the children within the same ward are correlated. Define a random variable

$$U_i(\psi) = \int_0^{T_i} \exp\{\psi A_i(t)\} ds$$

for model (1) and with an analogous definition of $U_i(\psi)$ for model (2). By assumption (3) we have at the true value $\psi = \psi_0$

$$U_i(\psi_0) \perp\!\!\!\perp R_i \mid Z_i, \tag{4}$$

i.e. for given Z_i the treatment-free survival is independent of the treatment assignment at the true value $\psi = \psi_0$. Condition (4) is the key for obtaining estimating equations for ψ. Robins and Tsiatis [5] derive an optimal and asymptotically normal estimator for ψ using a set of linear rank test statistics. We derive an estimator for ψ by making a *working assumption* that the treatment-free survival time U_i follows a proportional hazards model with respect to some baseline covariates. Since $U_i = U_i(\psi_0) = T_i(R_i = 0)$ this assumption can be checked using data from the placebo arm. By the working assumption we write the estimation model

$$\lambda_i(u) = \lambda_0(u) r_i\{\beta(\psi), \theta(\psi), Z_i, R_i\}, \tag{5}$$

which may take the form

$$\lambda_i(u) = \lambda_0(u) \exp\{\theta_1(\psi) R_i\} \text{ or} \tag{6}$$

$$\lambda_i(u) = \lambda_0(u) \exp\{\beta(\psi) Z_i + \theta_1(\psi) R_i\} \text{ or} \tag{7}$$

$$\lambda_i(u) = \lambda_0(u) \exp\{\beta(\psi) Z_i + \theta_1(\psi) R_i + \theta_2(\psi) R_i \times Z_i\}, \tag{8}$$

where $\lambda_0(u)$ denotes the baseline hazard of $U_i(\psi)$, $\beta(\psi)$ quantifies the dependence of the risk on the baseline covariates Z_i and $\theta(\psi)$ quantifies for the dependence of the risk on terms involving R_i. Model (6) corresponds to using the log-rank test statistic as an estimating equation. By (4) the relative risk function $r_i\{\beta(\psi), \theta(\psi), Z_i, R_i\}$ does not depend on terms involving R_i at the true $\psi = \psi_0$, or equivalently $\theta(\psi) = 0$ at $\psi = \psi_0$. Using the estimation model we construct a test for the null hypothesis $\psi = \psi_0 \Leftrightarrow \theta(\psi) = 0$ using a conditional score test statistic

$$Q(\psi) = S_\theta^T\{\hat{\beta}(\psi), \theta(\psi)\} I^{\theta\theta}\{\hat{\beta}(\psi), \theta(\psi)\} S_\theta^T\{\hat{\beta}(\psi), \theta(\psi)\} \tag{9}$$

where $S_\theta\{\beta(\psi),\theta(\psi)\} = \frac{\partial}{\partial\theta(\psi)} \log L\{\beta(\psi),\theta(\psi)\}$ denotes the score function of the partial likelihood $L\{\beta(\psi),\theta(\psi)\}$ for $U_i(\psi)$ with $S_\theta\{\hat\beta(\psi),\theta(\psi)\}$ evaluated at the value $\hat\beta(\psi)$ estimated subject to the constraint $\theta(\psi) = 0$, $I\{\beta(\psi),\theta(\psi)\} = -\frac{\partial^2}{\partial\beta(\psi)\partial\theta(\psi)} \log L\{\beta(\psi),\theta(\psi)\}$ and $I^{\theta\theta}\{\hat\beta(\psi),\theta(\psi)\}$ denotes the inverse of $I\{\theta(\psi),\theta(\psi)\} - I\{\beta(\psi),\theta(\psi)\}I^{-1}\{\beta(\psi),\beta(\psi)\}I\{\theta(\psi),\beta(\psi)\}$ evaluated at the value $\hat\beta(\psi)$ estimated subject to the constraint $\theta(\psi) = 0$. At the true value $\psi = \psi_0$ the score test statistics (9) has an asymptotic χ^2 distribution with p degrees of freedom, where p is the number of parameters in ψ. A point estimate for ψ is found as the point where $Q(\psi)$ is minimized. The $(1-\alpha)$ confidence set is defined as the set $\{\psi : Q(\psi) \le \chi^2_p(\alpha)\}$ where $\chi^2_p(\alpha)$ denotes the respective quantile of the χ^2 distribution.

3.2 Censoring by end of follow-up

With censored data $U_i(\psi)$ is not observable. Write $\eta_i = \exp(\psi)$ for model (1) and $\eta_i = \exp(\psi_1 + \psi_2 Z_{ij})$ for model (2). Define

$$C_i(\psi) = \begin{cases} \eta_i\, C_i, & \text{if } \eta_i < 1, \\ C_i, & \text{otherwise.} \end{cases}$$

Further define $X_i(\psi) = \min\{U_i(\psi), C_i(\psi)\}$ and $\Delta_i(\psi) = I\{U_i(\psi) = X_i(\psi)\}$. Because both $X_i(\psi)$ and $\Delta_i(\psi)$ are functions of baseline characteristics C_i, Z_i, and since $U_i(\psi)$ and the censoring due to end of follow-up do not depend on R_i then at the true value $\psi = \psi_0$

$$\{X_i(\psi), \Delta_i(\psi)\} \perp\!\!\!\perp R_i \mid Z_i. \tag{10}$$

For given value of ψ model (5) is fitted to the data $\{X_i(\psi), \Delta_i(\psi)\}$ and the score test statistic (9) is minimized with respect to ψ by repeating the procedure for a range of values of ψ.

3.3 Censoring by competing risks

When analysing for example time to lung cancer diagnosis (T_{2i}), some subjects die before reaching the endpoint. Here death (T_{1i}) is acting as a competing event and in the presence of competing events the variables $X_i(\psi)$ and $\Delta_i(\psi)$ are not always observable. In order to adjust our estimation method for competing events we use the approach proposed by Robins and coauthors [6, 29]. Define $X_i = \min(T_{1i}, T_{2i}, C_i)$. Let $\lambda_{T_1}(t \mid Z_i, L_i(t), R_i, X_i \ge t)$ denote the conditional hazard for the competing event at time t. An underlying untestable assumption needed is that at $\psi = \psi_0$ this hazard does not further depend on $U_i(\psi)$ when we condition on a sufficient number of baseline covariates Z_i and post-randomization covariates $L_i(t)$. For this assumption to hold at least approximately one needs to have a richly-parametrized model for $\lambda_{T_1}(t \mid Z_i, L_i(t), R_i, X_i \ge t)$ including all possible time-dependent and time-invariant covariates that might affect both T_{1i} and U_i. The conditional probability for subject i to survive until X_i is estimated by

$$\hat K_i = \hat K(X_i) = \exp\left\{ -\int_0^{X_i} \hat\lambda_{T_1}(t \mid Z_i, L_i(t), R_i, X_i \ge t)\, ds \right\}$$

using for example a Cox model with time-dependent covariates or a piecewise exponential hazard model. In the analysis we only use data $\{X_i, \Delta_i = I(X_i = T_{2i})\}$ from those who do not experience a competing event and for them $\hat{K}_i^{-1}$ is used as a weight in the estimation procedure. Heuristically this means that a person who did not die prior to end of follow-up or prior to lung cancer diagnosis accounts not only for himself but also for $\hat{K}_i^{-1} - 1$ other subjects that are similar to him but died before X_i. We need to modify the score test statistic (9) because $\hat{K}_i$ depends on all the data. This is accomplished by replacing $I\{\hat{\beta}(\psi), \theta(\psi)\}$ by a robust covariance matrix estimator [30] in the score test statistic (9). Robins [29] shows that this procedure leads to asymptotically conservative confidence intervals.

3.4 Modification for missing covariates in the accelerated failure time model

The above estimation methods are further generalized in Paper IV to the situation where some of the covariates in the structural model may have missing values. In Paper IV time of smoking cessation is observed only for those participants of the ATBC Study who report to have quit smoking before withdrawal. For subjects who withdraw the smoking status is unknown. To adjust the estimation procedure for missing information on smoking cessation we create $M = 3$ complete datasets as described in section 6 of Paper IV. We combine the tests statistics (9) from each of the $m = 1, \ldots, M$ complete data analyses for a given value ψ using two alternative tests. For the mth complete dataset let $S_{\theta,m} = S_\theta\{\hat{\beta}(\psi), \theta(\psi)\}$, $Q_m = Q(\psi)$ and $I_m^{\theta\theta} = I^{\theta\theta}\{\hat{\beta}(\psi), \theta(\psi)\}$. The first test statistic by Li et al. [31] is defined as

$$D(\psi) = \frac{\bar{S}_\theta^T \bar{U}^{-1} \bar{S}_\theta}{p(1+r)}$$

where $\bar{S}_\theta = M^{-1} \sum S_{\theta,m}$, $\bar{U} = M^{-1} \sum (I_m^{\theta\theta})^{-1}$, $r = p^{-1}(1 + M^{-1})\text{tr}(B\bar{U}^{-1})$ and $B = (M-1)^{-1} \sum (S_{\theta,m} - \bar{S}_\theta)(S_{\theta,m} - \bar{S}_\theta)^T$ with summation over m. Under the null the test statistic $D(\psi)$ has an F reference distribution with p and $\nu_1 = 4 + [p(M-1) - 4][1 + r^{-1}(1 - \frac{2}{p(M-1)})]^2$ degrees of freedom [31]. Another test statistic proposed by Rubin [32] combines directly the test statistics (9) and is defined as

$$\hat{\bar{D}}(\psi) = \frac{p^{-1}\bar{Q} - \frac{M-1}{M+1}\hat{r}}{1 + \hat{r}}$$

where $\bar{Q} = M^{-1} \sum Q_m$, $\hat{r} = \frac{(1+M^{-1})s^2}{2\bar{Q} + \max(0, 4\bar{Q}^2 - 2ps^2)}$ and $s^2 = (1-M)^{-1} \sum (Q_m - \bar{Q})^2$. Under the null $\hat{\bar{D}}(\psi)$ has an F reference distribution with p and $\nu_2 = 2^{-1}(1 + p^{-1})(M-1)(1 + \hat{r}^{-1})^2$ degrees of freedom. A point estimate for ψ is found as the value which minimizes $D(\psi)$ or $\hat{\bar{D}}(\psi)$ and the joint $(1-\alpha)$ confidence region is defined as the set $\{\psi : D(\psi) \leq F_{p,\nu_1}(\alpha)\}$ or $\{\psi : \hat{\bar{D}}(\psi) \leq F_{p,\nu_2}(\alpha)\}$ where $F_{p,\nu_1}(\alpha)$ and $F_{p,\nu_2}(\alpha)$ denote the $(1-\alpha)$ percentage points of the respective F reference distributions.

3.5 Previous applications of the accelerated failure time models

There exist several applications of the accelerated failure time models in the context of randomized and observational studies. Robins et al. [6] estimate the effect of a post-randomization prophylaxis treatment on the survival of AIDS patients in an observational study. There is no physical randomization in an observational study and for estimation purposes one needs to model explicitly the evolution of the treatment received over time. Robins et al. [6] model the hazard of initiation of the prophylaxis treatment at time t using a richly-parametrized time-dependent Cox model which conditions on a large number of time-dependent and time-invariant covariates. Assuming that the Cox model is correctly specified the inferences rely on the assumption that the hazard does not futher depend on the treatment-free survival time. This assumption is referred to as "the assumption of no unmeasured confounders". Further applications of accelerated failure time models in an observational setting can be found in Mark and Robins [33], Joffe et al. [34], Keiding et al. [35] and Witteman et al. [36].

The first use of accelerated failure time models in the context of a randomized clinical trial with non-compliance was by Mark and Robins [37], who estimated the effect of smoking cessation on overall survival in the Multiple Risk Factor Intervention Trial. Robins and Greenland [27] first used a two-parameter accelerated failure time model when estimating the direct effect of AZT treatment and the effect of a post-randomized prophylaxis treatment on survival in a randomized study with AIDS patients. They used two linear rank tests, the log-rank test and the Prentice-Wilcoxon test, to obtain the estimating equations for the two parameters of the accelerated failure time model. White and Goetghebeur [38] also used a two-parameter accelerated failure time model when studying the effect of an anti-hypertensive treatment on survival. They used the model, however as a tool for a sensitivity analysis when examining how post-randomized treatment changes affected the intention-to-treat estimates. White et al. [39] used a two-parameter model with the same estimating equations as in Robins and Greenland [27].

4 Summaries of the original papers

Each paper of this thesis focusses on a particular application aimed at answering specific subject matter research questions. The papers and the datasets of each application are summarized below.

4.1 Paper I. Correcting for non-compliance in randomized trials: An application to the ATBC Study

The application of Paper I arises from the ATBC Study [10]. Briefly, the ATBC Study was a randomized, double-blinded, placebo-controlled chemoprevention trial conducted in 14 adjoining areas of Finland between 1985 and 1993. The objectives of the study were to evaluate the effects of alpha-tocopherol and beta-carotene supplementations on lung cancer incidence, on incidence of

other major cancers and on total mortality in a cohort of 29133 middle aged male smokers being at high risk for lung cancer. The participants were recruited between 1985 and 1988. The scheduled end of the active intervention was April 30, 1993. 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, combination of them, or placebo. At the time of randomization the censoring time C_i was fixed and known for each participant and defined as the difference between April 30, 1993 and the date of randomization. Paper I considers only two randomized groups, namely the group of 14560 men who were assigned to receive beta-carotene (BC group) and 14573 men assigned not to receive beta-carotene (NOBC group). Note that the BC group includes combination treatment as well as beta-carotene treatment alone. Similarly, the NOBC group includes placebo treatment and alpha-tocopherol treatment alone. At baseline the participants were interviewed and details of their demographic variables, their medical, smoking, dietary and occupational history were recorded. Levels of alpha-tocopherol and beta-carotene in serum were also measured from blood samples drawn at baseline. During follow-up three annual visits to the local field center were made, during which the men were asked about their health status and smoking habits since the last visit. At the follow-up visits unused study capsules from the previous period were returned, counted and recorded, and a new pack 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 supplementation and were considered as withdrawals. Endpoint information was received from national registers regardless of the withdrawal status of the participants. Specifically, deaths were confirmed via the Central Population Register and lung cancers were identified through the Finnish Cancer Registry and further reviewed by the Clinical Review Commitee for confirmation and staging.

The intention-to-treat analyses suggested a harmful effect of beta-carotene on both lung cancer incidence and total mortality. The relative risk estimate was 1.08 (95 % CI from 1.01 to 1.16) for total mortality [11]. This is an estimate of the effectiveness. However 25 % of the men withdrew from the BC group for reasons other than death and time after withdrawal in the BC group accounted for about 15.1 % of the total follow-up time in that arm (Table IV of Paper I). Thus non-compliance caused by withdrawal from the BC arm is quite substantial and the estimate of effectiveness is attenuated compared to an estimate of efficacy. It is thus of interest to estimate the effect of beta-carotene supplementation actually received on overall survival.

Paper I compares three different approaches for the analysis: the intention-to-treat (ITT) approach based on a Cox proportional hazards model with the treatment assignment R_i as a covariate, the as-treated (AT) approach based a Cox proportional hazards model with the treatment received $A_i(t)$ as a time-dependent covariate and the g-estimation (GE) approach using a one-parameter accelerated failure time model with a log-rank test statistic as an estimating equation. The GE approach amounts to using the estimation method of section 3.1 with (6) as the estimation model. A simulation study performed under different non-compliance settings indeed showed that the ITT estimate is attenuated towards null due to non-compliance (Table III and Figures 1 and 2 of Paper I). It was further shown that the AT approach gives valid estimates only if (i) U_i follows an expo-

nential distribution and (ii) non-compliance is independent of U_i. The GE approach performs well in all non-compliance settings.

The analysis of the ATBC Study suggests that non-compliance is related to the treatment-free survival time through measured baseline covariates and possibly also through other unmeasured covariates, which induces non-ignorable non-compliance (Table V of Paper I). Thus the AT approach is inappropriate. Based on the GE approach the survival time is estimated to be 7.4 % shorter if one would always be on beta-carotene supplementation as opposed to never being on beta-carotene supplementation (Table VI of Paper I). Comparing this with the accelerated failure time ITT estimate of 5.4 % reduction demonstrates the attenuation.

4.2 Paper II. Effect modification in a randomized trial under nonignorable noncompliance: An application to the ATBC Study

The estimation method described in section 3.1 is developed in Paper II. The estimation model (7) which conditions on one baseline covariate is compared in simulations with Robin's g-estimation approach that uses (6) as the estimation model. Simulations show a substantial increase in precision with the former method (Table 1 of Paper II) and also that the method performs well under the null even if the estimation model miss-specified (Table 3 of Paper II). Paper II also provides simulation based techniques for evaluating the adequacy of the accelerated failure time model. The simulations suggest that when the correct model is (2) and the analysis is done without the interaction term using model (1) the method may detect such a model miss-specification (Figure 1 of Paper II). Miss-specification of the functional form of the covariate in the model (2) was, however, not detected.

The aim of the application of Paper II is to estimate the effect of beta-carotene actually received on total mortality in the ATBC Study and to quantify how effect modification by baseline covariates such as age, smoking, alcohol consumption and serum beta-carotene level. This paper is the first to address the estimation of an interaction in a real application. The analysis of Paper II is restricted to those 27025 men who did not have missing values on covariates of interest (Table 4 of Paper II). It was found that none of the studied interactions was statistically significant but all of them had the same signs as in the stratified analysis (Figure 3 and Table 5 of Paper II). It was estimated that for a person who smokes 10 cigarettes per day more than on average the lifespan would be 12.2 % shorter if one would always reveive beta-carotene supplementation than if one would never reveive beta-carotene supplementation. Similarly for a person with the daily alcohol comsumption 10 g/day more than on average the survival time would be 8.7 % shorter if always on beta-carotene supplementation than if never on beta-carotene supplementation. With increasing levels for the baseline serum beta-carotene level the harmful effect of beta-carotene supplementation decreased.

4.3 Paper III. Vitamin A and infant mortality: Beyond intention-to-treat in a randomized trial

In Paper III we investigate the effect of one oral dose of vitamin A on subsequent 4 month mortality in children under 6 months of age in a randomized, double-blind, placebo-controlled community trial in Nepal with partially unobserved non-compliance. Between September 1989 and December 1991, 261 *wards* in 29 contiguous village development areas in the rural plains of Nepal were randomly allocated either to receive an oral dose of vitamin A or placebo every 4 months [12]. Households in the wards were visited every 4 months with a maximum of 6 visits. Starting from the second visit, infants born since the previous visit were enrolled and dosed. In total 9178 newborns entered the study over the 5 visits of interest. At enrollment social economic information like the caste to which the children belong and the number of radios owned by their household were recorded. Arm circumference was measured at each visit and morbidity in the week preceding dosing was reported in terms of number of days the child had had diarrhea, fever and cough. We censored follow-up time at 4 months after the first dose was scheduled and before a second dose was possibly given. There is no loss-to-followup, and hence the censoring time is known at entry for each child. All children did not receive their vitamin A or placebo capsule: approximately 85 % were dosed immediately by the field staff and about 2 % did never receive their intended dose. For about 13 % of the children capsules were left at home with their parents and we do not know whether the capsule was taken or not. For these 13 % the precise compliance status is thus unknown. Table 1 of Paper III shows the number of children entering the study at visits 2 to 6 and in parentheses the number of those who died within 4 months stratified by treatment group and compliance category. Within the defined follow-up time, 173 of the 9178 children died (1.9 %), 77 in the placebo group and 96 in the vitamin A group.

The data of the trial offer two new challenges for the accelerated failure time model approach: (i) ward-based randomization induces correlation between survival outcomes within the same ward and (ii) the actual receipt of vitamin A dose is not always known. For (i) we use a robust version of the estimation method of Paper II that replaces $I\{\hat{\beta}(\psi), \theta(\psi)\}$ by a robust covariance matrix estimator in the score test statistic (9). For (ii) we use two different definitions for compliance we perform a sensitivity analysis which captures the boundaries for the estimated parameters. In the analysis it was found that the effect of vitamin A was beneficial in the beginning of the trial but towards the end of the trial there was a reversal of this effect (Tables 4 and 6 of Paper III).

4.4 Paper IV. Joint estimation of effects of beta-carotene supplementation and smoking cessation on overall survival and on time to lung cancer diagnosis in the ATBC Study

In Paper IV we estimate the effects of beta-carotene supplementation, smoking cessation and their interaction on overall survival and on time to lung cancer diagnosis in the ATBC Study. About

21 % of the men in the ATBC Study stopped smoking during the active follow-up period and a naive analysis of the data suggested that the harmful effect of beta-carotene supplementation on total mortality and on lung cancer incidence was more pronounced among quitters (Tables 1 and 2 of Paper IV). The naive analysis is subject to selection bias and should not be interpreted as evidence for an interaction. In order to evaluate the possible interaction we use accelerated failure time models where time to smoking cessation has a similar role as non-compliance because it is a post-randomization covariate. An additional challenge for the analysis arises because the smoking status after dropout is unknown for those men who were smokers at the time of dropout (Table 3 of Paper IV). Based on the work of Paper II we propose an extention to the inference prodecure that accounts for the unknown smoking status using a multiple imputation scheme. This was described in section 3.4. In the analysis of time to lung cancer diagnosis we adjust for censoring due to competing risk by death as described in section 3.3.

In the analysis of total mortality in the ATBC Study data we compared the performance of the two test statistics $D(\psi)$ and $\hat{\hat{D}}(\psi)$ which gave practically the same answers (Figure 1 of Paper IV). Because $\hat{\hat{D}}(\psi)$ requires less space for storage it was used in subsequent analyses. We found no evidence of an effect of smoking cessation or an interaction on either overall survival or on time to lung cancer diagnosis. A sensitivity analysis showed the effect of beta-carotene supplementation was harmful for both endpoints for a range of values for the effect of smoking cessation (Figures 3 and 4 of Paper IV). Under the assumption of no effect of smoking cessation the survival time was estimated to be 4.9 % (0.5 % to 9.1 %) shorter in the beta-carotene supplementation arm than in the arm with no beta-carotene supplementation (Table 4 of Paper IV). Similarly, the effect of beta-carotene supplementation was estimated to shorten the time to lung cancer diagnosis by 10.4 % (3.4 % to 16.9 %) (Table 5 of Paper IV).

5 Concluding remarks

Compliance issues have recently gained much attention in the medical statistics literature. An example of the increasing interest among researchers is the one-day compliance symposium arranged at Limburg University in 1995 which was publised in Statistics in Medicine [40]. While the analysis by intention-to-treat remains the gold standard for comparing different treatment strategies in a clinical trial setting, information on post-randomization covariates such as compliance can give useful knowledge about the efficacy of the treatment. The estimation of the efficacy is not possible without additional non-identifiable assumptions. The perhaps strongest assumption made throughout this thesis is that the treatment effect can be captured by the accelerated failure time model. In Papers I and II we provide some graphical and simulation-based techniques for evaluating whether an interaction term should be included in the accelerated failure time model, but otherwise we need to rely on the correctness of the specified accelerated failure time model. With competing risks a further assumption of "no unmeasured confounders" is introduced, which cannot be empirically tested. As pointed out by Dunn [41] an underlying assumption that should be recognized is that

compliance is assumed to be error-free. It is plausible that the actual receipt of treatment is measured with error especially in the case where compliance is assessed through capsule count as in the ATBC Study. This will cause the estimates to be biased and modification of the accelerated failure time modelling approach to allow for measurement error is thus motivated.

In Paper II we develop an estimation method which can easily be used with a multi-parameter structural accelerated failure time model. The efficiency of the proposed estimation procedure was demonstrated in simulations in Paper II. The power of the estimation method is still rather poor. Robins and Tsiatis [5] describe a class of estimators which includes an optimal estimator whose asymptotic variance attains a semi-parametric efficiency bound. They discuss an adaptive procedure to find an estimator whose asymptotic efficiency may be close to that of the optimal estimator. The practical impementation of this procedure would, however, be tedious and the magnitude of the gain in efficiency is likely to be of little practical consequence. Another way of increasing efficiency is to turn the randomized trial into an observational study where one explicitly models the active treatment process after randomization. Under the assumption of "no unmeasured confounders" one can obtain estimating equations for the treatment effect parameter. One practical example of this is the observational analysis by Mark and Robins [33], which shows a significant improvement in efficiency over the previous analysis which used a log-rank test statistic for estimation [37]. The drawback of this approach is, however, that the assumption of "no unmeasured confounders" cannot be empirically tested and thus may result in biased estimation.

Loeys and Goetghebeur [42] suggested a modification to our estimation procedure by estimating $\beta(\psi)$ from the estimation model (5) using data only from the placebo arm. They obtained a closed form estimator for the structural parameter in the binary compliance case by assuming exponentially distributed treatment-free survival time. However, in the more complex situation of continuous compliance the structural parameter estimator does not exist in closed form.

The analyses from the ATBC Study in Paper I and II suggest that the effect of beta-carotene supplementation is more harmful when the analyses are adjusted for non-compliance. This is a consequence of models (1) and (2) which assume a constant acceleration rate over time. The accelerated failure time model can be modified to accomodate different dose-response mechanisms, for instance by using the cumulative dose process $\int_0^t A_i(u)du$ instead of $A_i(t)$. The models (1) and (2) also assume that after withdrawal there is no treatment effect. Again, a more flexible model that allows the treatment to have an effect on survival also after withdrawal could be used. While the power of the proposed estimation methods remains to be improved this thesis provides estimation tools that easily extend to more complex models of treatment action when addressing the effect treatment received on a survival outcome in a randomized placebo-controlled study with non-ignorable non-compliance.

References

[1] A. L. Cochrane. *Effectiveness and efficacy: Random reflections on health services*. Nuffield Provincial Hospitals Trust, London, 1972.

[2] D. Schwartz and J. Lellouch. Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Diseases*, 20:637–648, 1967.

[3] R. J. A. Little and D. B. Rubin. *Statistical analysis with missing data*. John Wiley & Sons, New York, 1987.

[4] J. M. Robins. 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*, pages 113–159. NCHSR, U.S. Public Health Service, 1989.

[5] J. M. Robins and A. A. Tsiatis. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics, Theory and Methods*, 20:2609–2631, 1991.

[6] J. M. Robins, D. Blevins, G. Ritter, and M. Wulfsohn. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia (PCP) on the survival of AIDS patients. *Epidemiology*, 3:319–336, 1992.

[7] J. M. Robins. Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, 79:321–334, 1992.

[8] J. M. Robins. Structural nested failure time models. In P. Armitage and T. E. Colton, editors, *Encyclopedia of Biostatistics*, pages 4372–4389. John Wiley & Sons, Chichester, 1998.

[9] D. R. Cox and D. Oakes. *Analysis of survival data*. Chapman & Hall, London, 1984.

[10] 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.

[11] 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.

[12] K. P. West, J. Katz, S. R. Sherestha, S. C. LeClerq, S. K. Khatry, E. K. Pradhan, R. Adhikari, L. S.-F. Wu, R. P. Pokhrel, and A. Sommer. Mortality of infants under 6 mo of age supplemented with vitamin A: a randomized, double-masked trial in Nepal. *American Journal of Clinical Nutrition*, 62:143–148, 1995.

[13] D. B. Rubin. Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.

[14] D. B. Rubin. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, 6:34–58, 1978.

[15] P. W. Holland. Statistics and causal inference. *Journal of the Americal Statistical Association*, 81:945–960, 1986.

[16] I. Tarwotjo, A. Sommer, K. P. West, E. Djunaedi, L. Mele, B. Hawkins, and the ACEH Study Group. Influence of participation on mortality in a randomized trial of vitamin A prophylaxis. *American Journal of Clinical Nutrition*, 45:1466–1471, 1987.

[17] A. Sommer and S. Zeger. On estimating efficacy from clinical trials. *Statistics in Medicine*, 10:45–52, 1991.

[18] J. Cuzick, R. Edwards, and N. Segnan. Adjusting for non-compliance and contamination in randomized clinical trials. *Statistics in Medicine*, 16:1017–1029, 1997.

[19] E. Goethgebeur and G. Molenberghs. Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *Journal of the Americal Statistical Association*, 91:928–934, 1996.

[20] J. D. Angrist, G. W. Imbens, and D. B. Rubin. Identification of causal effects using instrumental variables (with comments). *Journal of the Americal Statistical Association*, 91:444–472, 1996.

[21] G. W. Imbens and D. B. Rubin. Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, 25:305–327, 1997.

[22] B. Efron and D. Feldman. Compliance as an explanatory variable (with discussion). *Journal of the Americal Statistical Association*, 86:9–26, 1991.

[23] J. M. Albert and D. L. DeMets. On a model-based approach to estimating efficacy in clinical trials. *Statistics in Medicine*, 13:2323–2335, 1994.

[24] E. T. Goethgebeur and S. H. Shapiro. Analysing non-compliance in clinical trials: Ethical imperative or mission impossible? *Statistics in Medicine*, 15:2813–2826, 1996.

[25] E. Goethgebeur and K. Lapp. The effect of treatment compliance in a placebo-controlled trial: regression with unpaired data. *Applied Statistics*, 46:351–364, 1997.

[26] J. M. Robins. Correcting for noncompliance using structural nested mean models. *Communications in Statistics, Theory and Methods*, 23:2379–2412, 1994.

[27] J. M. Robins and S. Greenland. 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 Americal Statistical Association*, 89:737–749, 1994.

[28] C. E. Frangakis and D. B. Rubin. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-compliance and subsequent missing outcomes. *Biometrika*, 86:365–379, 1999.

[29] J. M. Robins. Analytic methods for estimating HIV treatment and cofactor effects. In *Methodological issues in AIDS mental health research*, pages 213–290. NCHSR, U.S. Public Health Service, New York: Plenum Press, 1993.

[30] L. J. Wei, D. Y. Lin, and L. Weissfeld. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the Americal Statistical Association*, 84:1065–1073, 1989.

[31] K. H. Li, T. E. Raghunathan, and D. B. Rubin. Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the Americal Statistical Association*, 86:1065–1073, 1991.

[32] D. B. Rubin. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, 1987.

[33] S. D. Mark and J. M. Robins. Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model. *Statistics in Medicine*, 12:1605–1628, 1993.

[34] M. M. Joffe, D. R. Hoover, L. P. Jacobson, L. Kingsley, J. S. Chmiel, B. R. Visscher, and J. M. Robins. Estimating the effect of zidovudine on Kaposi's sarcoma from observational data using a rank preserving structural failure-time model. *Statistics in Medicine*, 17:1073–1102, 1998.

[35] N. Keiding, M. Filiberti, S. Esbjerg, J. M. Robins, and N. Jacobsen. The graft versus leukemia effect after bone marrow transplantation: A case study using structural nested failure time models. *Biometrics*, 55:23–28, 1999.

[36] J. M. Witteman, R. B. D'Agostino, T. Stijnen, W. B. Kannel, J. C. Cobb, M. A. J. de Ridder, A. Hofman, and J. M. Robins. G-estimation of causal effects: Isolated systolic hypertension and cardiovascular death in the Framingham Heart Study. *American Journal of Epidemiology*, 148:390–401, 1998.

[37] S. D. Mark and J. M. Robins. 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.

[38] I. R. White and E. J. T. Goetghebeur. Clinical trials comparing two treatment policies: which aspects of the treatment policies make a difference. *Statistics in Medicine*, 17:319–339, 1998.

[39] I. R. White, A. G. Babiker, S. Walker, and J. H. Darbyshire. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. *Statistics in Medicine*, 18:2617–2634, 1999.

[40] Limburg Compliance Symposium. *Statistics in Medicine*, 17:249–389, 1998.

[41] G. Dunn. The problem of measurement error in modelling the effect of compliance in a randomized trial. *Statistics in Medicine*, 18:2863–2877, 1999.

[42] T. Loeys and E. Goetghebeur. A note on binary non-compliance and exponential survival models in the context of randomized trials. Technical report, Department of Applied Mathematics and Informatics, University of Ghent, 1999.