Published in final edited form as:
Stat Med. 2011 March 30; 30(7): 709–717. doi:10.1002/sim.4131.

A method to estimate treatment efficacy among latent subgroups of a randomized clinical trial

Lily L. Altstein[a,*,†], **Gang Li**[a], and **Robert M. Elashoff**[a,b]

[a]Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90095-1772, U.S.A

[b]Department of Biomathematics, UCLA Medical Center, Box 951766, Los Angeles, CA 90095-1766, U.S.A

Abstract

Subgroup analysis arises in clinical trials research when we wish to estimate a treatment effect on a specific subgroup of the population distinguished by baseline characteristics. Many trial designs induce latent subgroups such that subgroup membership is observable in one arm of the trial and unidentified in the other. This occurs, for example, in oncology trials when a biopsy or dissection is performed only on subjects randomized to active treatment. We discuss a general framework to estimate a biological treatment effect on the latent subgroup of interest when the survival outcome is right-censored and can be appropriately modelled as a parametric function of covariate effects. Our framework builds on the application of instrumental variables methods to all-or-none treatment noncompliance. We derive a computational method to estimate model parameters via the EM algorithm and provide guidance on its implementation in standard software packages. The research is illustrated through an analysis of a seminal melanoma trial that proposed a new standard of care for the disease and involved a biopsy that is available only on patients in the treatment arm.

Keywords

survival analysis; accelerated failure time model; treatment noncompliance; mixture model; EM algorithm

1. Introduction

Clinical trials often aim to estimate a treatment effect that pertains to a specific subgroup of the patient population defined by baseline characteristics. The analysis becomes complicated when the subgroup of interest is latent and cannot be identified based on observed data. This occurs in the Multicenter Selective Lymphadenectomy Trial I (MSLT-I), an ongoing trial comparing two standards of care for primary cutaneous melanoma in which a sentinel-node biopsy distinguishing two subgroups of patients is available only on patients randomized to the treatment arm [1, 2]. Treatment patients whose biopsies are positive for nodal metastases (‘node-positive’) receive an immediate elective lymphadenectomy, a surgical procedure to remove the lymph node, while those absent metastases (‘node-negative’) are not eligible for the surgery and receive clinical observation, which is the control regimen and represented the prevailing course of treatment at the time the trial began in 1994. The primary objective

[*]Correspondence to: Lily L. Altstein, Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90095-1772, U.S.A.

[†]laltstein@ucla.edu

of MSLT-I is to evaluate the effectiveness of the new standard of care for melanoma; it is assessed by the traditional intention-to-treat (ITT) analysis comparing the two randomization arms. A secondary objective aims to estimate the effect of lymph node dissection on the node-positive subgroup. We term this estimand *biological efficacy* to distinguish it from the ITT estimand, and it is critical to this high-risk cohort whose mortality is over 30 per cent. Estimation of biological efficacy in MSLT-I involves latent subgroup analysis because the biopsy constitutes part of the experimental treatment strategy and hence is not administered to the control patients.

Latent subgroup analysis arises in the application of the instrumental variables (IV) framework to all-or-none treatment noncompliance [3, 4]. Subgroups in this setting are defined according to baseline potential compliance behaviors and are partially observed after randomization. This is analogous to MSLT-I, where nodal status is a pre-randomization covariate that is subsequently identified in the treatment arm through the sentinel-node biopsy. Table I illustrates the duality between MSLT-I and a trial with noncompliance when controls are denied access to treatment. In the presence of noncompliance, the ITT analysis assesses *programmatic effectiveness* but is biased for the biological or causal effect of treatment [5]. Furthermore, analyses based on treatment receipt or adherence to protocol are confounded by differences between patients across the compliance strata. This framework posits that the efficacy parameter of interest is defined in the class of *compliers*—individuals who accept whichever treatment is assigned to them—because this is the only class whose treatment receipt is determined by the randomization assignment, thus preserving the use of randomization as the instrument for causal inference within the compliant class. Comprehensive summaries of this approach can be found in Palmgren and Goetghebeur [6] and Dunn *et al.* [7].

The IV framework for all-or-none noncompliance was extended to censored data by Frangakis and Rubin [8]; however, the nonparametric estimator of the survival distribution function that results from this approach neither does not possess desirable statistical properties such as monotonicity [9, 10], nor does it accommodate covariates. These issues lead to several semiparametric developments that estimate biological efficacy among the subgroup of compliers with the aid of proportional hazards assumptions. Loeys and Goetghebeur [11] use isotonic regression to enforce monotonicity in a causal proportional hazards model but did not derive the asymptotic properties of their estimator. Cuzick *et al.* [12] proposed a Cox-type model that assumes proportionality of both the efficacy effect and all baseline hazard functions of the various compliance classes; a more complicated estimation procedure is required to fit covariates and interactions. The only example of a fully parametric regression model that we are aware of is in Follmann [13], which is a propensity score method that requires extrapolation of a pseudolikelihood estimate to unobserved values of the compliance covariate. Many accelerated failure time (AFT) models for survival analysis in the presence of treatment noncompliance exist in the structural models literature [14–16], but they do not pertain to subgroup analysis.

In this paper, we develop the framework for latent subgroup survival analysis when the survival distributions can be appropriately modelled as parametric functions of covariate effects. We derive a computational method that relies on the EM algorithm [17] for parameter estimation in an AFT mixture model. The EM algorithm has been applied to estimate efficacy in noncompliance problems in non-survival settings [18, 19] and produces maximum likelihood estimates with the desirable properties of consistency and asymptotic normality. Our method has the paramount advantage of being easily implemented in software such as SAS® or R [20] using existing routines. It readily incorporates covariates and their interactions with treatment and subgroup and allows flexibility in model specification sufficient to many applications. The framework and computational method are described in Sections 2 and 3, respectively, followed by a simulation study to validate the performance of the method in Section 4. We illustrate the method in Section 5 by estimating the effect of immediate lymphadenectomy on two disease-free survival (DFS) endpoints in the subgroup of MSLT-I patients with sentinel-node metastases.

2. The latent subgroup framework

Latent subgroup analysis aims to estimate a biological treatment effect that applies to a specific group of *treatable* patients who stand to benefit therapeutically from the treatment, for example MSLT-I patients with sentinel-node metastases. The remaining set of *untreatable* patients often receive the same care under both randomization assignments, as in MSLT-I, although this is not a requirement of our framework. Subgroup status is revealed through randomization to active treatment but is unidentified in control based on observed data, resulting in the statistical problem of inferring the subgroup of treatable controls.

2.1. Notation

The randomly right-censored survival times and censoring indicators are denoted (X_i, δ_i), i = 1, …,n, where X_i is the minimum of an event time T_i and an independent censoring time C_i: $X_i = T_i \wedge C_i$, $\delta_i = I(T_i \le C_i)$. Patients are randomized to the treatment ($R_i = 1$) or control ($R_i = 0$) arm, and additional data on q covariates may be measured at baseline, $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iq})'$. Let G_i be an indicator of membership to the treatable subgroup (for simplicity, we consider two subgroups). The observed data consists of (X_i, δ_i, R_i = 1,$\mathbf{Z}_i$, G_i) for an individual randomized to treatment and (X_i, δ_i, R_i = 0,$\mathbf{Z}_i$) for an individual randomized to control; we will refer to the aggregate set of observed data as $\mathbf{X_{obs}}$. We denote the marginal survival function as $S(t|\mathbf{Z}) = P(T > t|\mathbf{Z})$, and those conditional on G and R by $S_{gr}(t|\mathbf{Z}) = P(T > t|\mathbf{Z}, G = g, R = r)$ for g = 0,1 and r = 0,1; all are conditional on the fixed covariate vector $\mathbf{Z}$.

2.2. Model and assumptions

The AFT mixture model for latent subgroup analysis conditions on subgroup status G:

$$\begin{aligned} p=P(G_i=1), \\ \log(T_i|\mathbf{Z}_i, R_i, G_i=g)=\psi_g R_i+\mathbf{Z}_i'\beta_g+\sigma_g\varepsilon_g, \end{aligned} \quad (1)$$

i = 1, …,n, where p represents the population proportion of treatable subjects, and ψ_g and $\boldsymbol{\beta}_g$ the biological efficacy and other covariate effects in the gth subgroup, respectively, g = 0, 1. The parametric AFT model assumes that the survival time in each subgroup is a log-linear function of the covariates and a known residual error term, ε_g, which can differ over subgroup in both family and scale (σ_g). We further assume that the censoring mechanism is noninformative and independent of subgroup status. For simplicity, we have included the intercepts in $\boldsymbol{\beta}_g$. Shared covariate effects or residual error distributions across subgroups can be achieved through parameter constraints.

The notion of treatability translates to the *exclusion restriction*, an assumption formalized by Angrist *et al.* [3] that is fundamental to our framework. The exclusion restriction posits no direct effect of randomization on the response and hence that the outcome distribution of the untreatable subgroup is the same in both arms. In our model, it implies that $\psi_0 = 0$ and $S_{01}(t|\mathbf{Z}) = S_{00}(t|\mathbf{Z}) \equiv S_0(t|\mathbf{Z})$. This assumption is usually tenable when the untreatable subgroup receives the same treatment regardless of the randomization assignment; it is discussed further in the context of MSLT-I in Section 5. In addition to the exclusion restriction, our scenario satisfies the stable unit treatment value (SUTV), the random assignment, and the nonzero effect of randomization on treatment receipt assumptions of Angrist *et al.* [3]. Monotonicity is induced by design.

2.3. Model specification

Specification of the residual error distribution in an AFT model is traditionally performed through graphical analysis of transformations of the observed survival times and model residuals and through likelihood-ratio tests of nested models. Since the survival distributions in the two subgroups do not depend on the randomization assignment, these methods extend directly to our model when performed on the complete subgroup data from the treatment arm alone. Comparison of model-based survival distributions to Kaplan–Meier estimates yields post-hoc assessments of goodness-of-fit. Our model supports covariate by subgroup interactions by allowing separate model specifications in each subgroup, and maximum likelihood inference about individual and interaction effects is straightforward. Inclusion of covariates that are predictive of baseline response can increase precision for estimating efficacy, and covariate by subgroup interactions can be explored in the treatment arm.

3. Computational method

The likelihood corresponding to our model in (1) contains an observed mixture in the treatment arm, where subgroup status is known, and a latent mixture in the control arm over the treatable and untreatable subgroups: $S_0(t|\mathbf{Z})=pS_{10}(t|\mathbf{Z})+(1-p)S_0(t|\mathbf{Z})$. Thus the observed data likelihood for the parameter vector $\Theta = (p, \psi_1, \boldsymbol{\beta}_1, \sigma_1, \boldsymbol{\beta}_0, \sigma_0)$ is

$$
\begin{aligned}
L(\Theta|\mathbf{X_{obs}})= & \prod_{i:R_i=1}[pf_{11}(x_i|\mathbf{z}_i)^{\delta_i}S_{11}(x_i|\mathbf{z}_i)^{1-\delta_i}]^{g_i}[(1-p)f_0(x_i|\mathbf{z}_i)^{\delta_i}S_0(x_i|\mathbf{z}_i)^{1-\delta_i}]^{1-g_i}\\
& \prod_{i:R_i=0}[pf_{10}(x_i|\mathbf{z}_i)+(1-p)f_0(x_i|\mathbf{z}_i)^{\delta_i}[pS_{10}(x_i|\mathbf{z}_i)+(1-p)S_0(x_i|\mathbf{z}_i)]^{1-\delta_i},
\end{aligned}
\tag{2}
$$

where $f_{gr}(t) = -\mathrm{d}S_{gr}(t)$. It is not readily solved by traditional maximum likelihood routines that fit AFT models because of the mixture parameter embedded in the likelihood. The EM algorithm is a natural choice for parameter estimation when complete information on subgroup membership would allow factorization of the likelihood into separate components that are easy to maximize [4]. The complete data consists of the hypothetical data set in which subgroup status were known on all individuals: $\mathbf{X_{com}} = (\mathbf{X_{obs}}, g_i; i\text{: } R_i = 0)$. The corresponding likelihood is

$$
\begin{aligned}
L(\Theta|\mathbf{X_{com}})= & \prod_{i=1}^{n}\prod_{r\in\{0,1\}}[pf_{1r}(x_i|\mathbf{z}_i)^{\delta_i}S_{1r}(x_i|\mathbf{z}_i)^{1-\delta_i}]^{g_i}[(1-p)f_0(x_i|\mathbf{z}_i)^{\delta_i}S_0(x_i|\mathbf{z}_i)^{1-\delta_i}]^{1-g_i}\\
& =L(p|\mathbf{X_{com}})L(\psi_1,\beta_1,\sigma_1,\beta_0,\sigma_0|\mathbf{X_{com}}).
\end{aligned}
\tag{3}
$$

The log of this likelihood is linear in the unknown subgroup indicators in control, $(g_i; i\text{: } R_i = 0)$, and hence the E-step to compute its expectation conditional on the observed data and a current value of the parameter vector, $\Theta^{(m)}$, amounts to the imputation of the unknown g_i in control:

$$
\begin{aligned}
& E[G_i|\mathbf{X_{obs}},\Theta^{(m)}]=P(G_i=1|\mathbf{X_{obs}},\Theta^{(m)})\\
& =\delta_i\,P(G_i=1|T_i=x_i,C_i>x_i,\mathbf{z}_i,\Theta^{(m)})+(1-\delta_i)P(G_i=1|T_i>x_i,C_i=x_i,\mathbf{z}_i,\Theta^{(m)}).
\end{aligned}
$$

Under the assumptions that $C \perp T$ and $C \perp G$, we apply the Bayes Theorem to compute the expectation in censored and uncensored cases as follows. For $\delta_i = 1$,

$$P(G_i{=}1|T_i{=}x_i, C_i{>}x_i, \mathbf{z}_i, \Theta^{(m)}){=}\frac{P(T_i{=}x_i|G_i{=}1,\mathbf{z}_i,\Theta^{(m)})P(G_i{=}1|\Theta^{(m)})}{\sum_g P(T_i{=}x_i|G_i{=}g,\mathbf{z}_i,\Theta^{(m)})P(G_i{=}g|\Theta^{(m)})}$$
$$={\frac{p^{(m)}f_{10}(x_i|\mathbf{z}_i,\Theta^{(m)})}{p^{(m)}f_{10}(x_i|\mathbf{z}_i,\Theta^{(m)})+(1-p^{(m)})f_0(x_i|\mathbf{z}_i,\Theta^{(m)})}}.$$

Similarly, for $\delta_i = 0$,

$$P(G_i{=}1|T_i{>}x_i, C_i{=}x_i, \mathbf{z}_i, \Theta^{(m)}){=}\frac{P(T_i{>}x_i|G_i{=}1,\mathbf{z}_i,\Theta^{(m)})P(G_i{=}1|\Theta^{(m)})}{\sum_g P(T_i{>}x_i|G_i{=}g,\mathbf{z}_i,\Theta^{(m)})P(G_i{=}g|\Theta^{(m)})}$$
$$={\frac{p^{(m)}S_{10}(x_i|\mathbf{z}_i,\Theta^{(m)})}{p^{(m)}S_{10}(x_i|\mathbf{z}_i,\Theta^{(m)})+(1-p^{(m)})S_0(x_i|\mathbf{z}_i,\Theta^{(m)})}}.$$

The result of the *m*th E-step is the imputed variable:

$$g_i^{(m)}{=}\begin{cases} G_i & :R_i{=}1 \\ \delta_i\frac{p^{(m)}f_{10}(x_i|\mathbf{z}_i,\Theta^{(m)})}{p^{(m)}f_{10}(x_i|\mathbf{z}_i,\Theta^{(m)})+(1-p^{(m)})f_0(x_i|\mathbf{z}_i,\Theta^{(m)})} & \\ +(1-\delta_i)\frac{p^{(m)}S_{10}(x_i|\mathbf{z}_i,\Theta^{(m)})}{p^{(m)}S_{10}(x_i|\mathbf{z}_i,\Theta^{(m)})+(1-p^{(m)})S_0(x_i|\mathbf{z}_i,\Theta^{(m)})} & :R_i{=}0. \end{cases}$$

In the subsequent M-step, we compute

$$\Theta^{(m+1)}{=}\underset{\Theta}{\operatorname{argmax}}[L(p|\mathbf{X_{obs}}, g^{(m)})L(\psi_1, \beta_1, \sigma_1, \beta_0, \sigma_0|\mathbf{X_{obs}}, g^{(m)})],$$

which yields $p^{(m+1)}{=}\frac{1}{n}\sum_i^n g_i^{(m)}$ for the proportion of treatable subjects. If two separate AFT models are specified for the subgroups such that $L(\psi_1, \boldsymbol{\beta}_1, \sigma_1, \boldsymbol{\beta}_0, \sigma_0|\mathbf{X_{com}}) = L(\psi_1, \boldsymbol{\beta}_1, \sigma_1|\mathbf{X_{com}})L(\boldsymbol{\beta}_0, \sigma_0|\mathbf{X_{com}})$, the M-step for the remaining parameters can be performed by a weighted survival routine, available in most software packages, where the weights are given by $g_i^{(m)}$ and $1 - g_i^{(m)}$ for the treatable and untreatable subgroups, respectively. The E and M steps are iterated until a suitable convergence criterion is met, and the maximum likelihood estimate $\hat{\Theta}$ of Θ is the output of the last M-step.

In cases where we wish to constrain some of the parameters in order to reduce dimensionality, the M-step can be carried out through the PARAMEST macro [21] available in the SAS® software, which computes maximum likelihood estimates in arbitrary parametric AFT models. For example, to fit Weibull models to both subgroups with a common shape parameter ($\lambda = \sigma^{-1}$) we would specify the following hazard and survival functions as inputs to the macro at the *m*th step:

$$h^{(m)}(x_i|\mathbf{X_{obs}}, g_i^{(m)}){=}\lambda x_i^{\lambda-1}[\mathrm{e}^{(-\lambda(\psi_1 R_i+\mathbf{z}_i' \beta_1))}]^{g_i^{(m)}}[\mathrm{e}^{(-\lambda(\mathbf{z}_i' \beta_0))}]^{1-g_i^{(m)}}$$
$$S^{(m)}(x_i|\mathbf{X_{obs}}, g_i^{(m)}){=}[\mathrm{e}^{-(x_i\exp\{-(\psi_1 R_i+\mathbf{z}_i' \beta_1)\})^{\lambda}}]^{g_i^{(m)}}[\mathrm{e}^{-(x_i\exp\{-\mathbf{z}_i' \beta_0\})^{\lambda}}]^{1-g_i^{(m)}}.$$

This model represents a special case of proportionality of hazards across subgroup. Alternatively, shared parameter models can be fit in any software that performs weighted AFT regression using a data augmentation trick. A sample SAS® program illustrating the implementation of the model, including the shared parameter case, is available on the author's website at http://www.biostat.ucla.edu/Directory/Gli/personal/software.html.

The EM algorithm does not immediately provide the variance matrix of $\hat{\Theta}$. We have estimated this quantity in our simulations and the analysis of MSLT-I as the inverse of the observed Fisher Information:

$$\widehat{\mathrm{Var}}(\widehat{\Theta})=-\left[\sum_{i=1}^{n}\frac{\partial^2}{\partial\Theta^2}l(\Theta|\mathbf{X_{obs}})|_{\mathbf{\Theta}=\widehat{\mathbf{\Theta}}}\right]^{-1},$$

where the $l(\cdot)$ denotes a log-likelihood. One of the advantages of our method is that it provides maximum likelihood estimates whose asymptotic normality and consistency lead to a straightforward framework for inference about biological efficacy and other parameters.

4. Simulation study

We assessed the performance of the computational algorithm in two simulation studies. The first is designed to evaluate the general performance of the method under a variety of censoring rates and treatable subgroup proportions in both large (N = 1000) and small (N = 400) samples. Subjects were randomly assigned to treatment or control in a 50:50 ratio, and subgroup status was determined independently through Bernoulli samples. The survival times are Weibull-distributed conditional on the randomization assignment and subgroup membership with independent censoring times. The second simulation study was designed to emulate the rather extreme characteristics of MSLT-I—the censoring rate is greater than 70 per cent, and the observed proportion of node-positive patients is 15 per cent—under three putative efficacies: (ψ_1 = 0, 0.5, 1). The sample size is 1300, with 60 per cent allocated to the treatment arm, and the independent survival and censoring times are Weibull-distributed.

Table II(a)–(c) summarizes the results on the efficacy parameter ψ_1 from 1500 data sets simulated under each set of conditions. The coverage represents the proportion of time a 95 per cent normal confidence interval for ψ_1 contained the true value of the parameter, $\hat{\psi}_1$ is the average of efficacy estimates, the simulation variance is the empirical variance of the estimates $\hat{\psi}_1$, and the estimated variance is the average of the estimated variances $\widehat{\mathrm{Var}}(\widehat{\psi}_1)$. The results are consistent with our expectation for maximum likelihood estimation. In large samples (Table II(a)), the algorithm overestimates the variance of ψ_1 when the censoring rate is high, but the loss in power is negligible. The coverage on all other model parameters (β_0, β_1, σ_0, σ_1) ranged from 94 to 96 per cent. Unsurprisingly, the algorithm tends to perform better with respect to bias and coverage on the parameters corresponding to the majority population when the true proportion of treatable subjects approaches 0 or 1. The estimation procedure performs well in small sample sizes (Table II(b)), especially when the treatable proportion is high, as is typically the case in a noncompliance setting.

5. Analysis of MSLT-I

Beginning in 1994, MSLT-I randomized patients with invasive primary cutaneous melanoma to receive an experimental course of treatment or control in a 60:40 ratio. A wide excision of the primary melanoma was performed on all patients, after which the control patients received postoperative observation of regional lymph nodes and treatment patients received a sentinel-node biopsy with immediate lymphadenectomy if the biopsy was positive for nodal micrometastases. Control patients and node-negative treatment patients could elect to receive a lymphadenectomy if the cancer recurred. The third interim analysis represents the most current findings from the trial, and it focused on the ITT test of effectiveness of the new standard of care on all patients [22]. The results were negative in part, which is unsurprising because the majority (80 per cent) of patients are absent metastases and receive the same course of treatment (postoperative observation) regardless of the randomization assignment. Subgroup analysis of the effect of lymphadenectomy on node-positive patients is a secondary aim of the trial, to which end the investigators compared node-positive patients in treatment to control patients who later developed recurrence. This strategy produces a biased estimate of biological efficacy because there is not a one-to-one correspondence between positive biopsy at randomization and eventual development of a clinically evident recurrence.

We apply the method to reanalyze two endpoints from the 2006 interim analysis: disease-free survival (DFS), defined as time until a clinically detectable recurrence, and distant-disease-free survival (DDFS), defined as time until recurrence at a distant site or melanoma death without distant recurrence. No melanoma deaths occurred without some form of recurrence: nodal, local or distant. The latter are the most baneful; hence, DDFS is considered more relevant than overall DFS. Diagnostic plots of data from the treatment arm supported the fit of the log-linear Weibull model with a common scale parameter across subgroup for both DFS and DDFS. We note further that in the context of MSLT-I, the exclusion restriction amounts to the assumption that sentinel-node biopsy does not alter the prognosis of node-negative patients. This is widely believed to be true among clinicians owing to the minimal invasiveness of the biopsy [2, 22].

The primary model to estimate the efficacy of lymphadenectomy on each endpoint is $\log(T_i \mid G_i = g)= \beta_g + \psi_g R_i + \sigma\varepsilon$, $g = 0,1$, where $\psi_0 = 0$ and $R_i = 1$ if the patient was randomized to the treatment arm and zero otherwise. The treatment significantly prolongs both DFS and DDFS in node-positive patients (Table III(c): $\hat{\psi}_1 = 1.593$, p-value=2.4E–13 and $\hat{\psi}_1 = 0.937$, p-value = 0.0247, respectively). Node-negative patients have the highest rates of DFS and DDFS (p-values all < 0.0001 when compared to the node-positive subgroup assigned biopsy). The estimate of p, the population proportion of node-positive subjects, is similar in both regressions: approximately 0.16 with a 95 per cent confidence interval of (0.14, 0.18). For comparison, we also fit standard Weibull AFT models to estimate the ITT estimand of effectiveness on each endpoint (Table III(b)); the test is significant for DFS but not for DDFS.

We fit a follow-up regression to investigate the effects of Breslow thickness, which measures the thickness of the melanoma, presence of ulceration, and location of the melanoma on the trunk as opposed to the head, neck or extremities on DDFS. These covariates were selected on the basis of prior research showing their significance to disease progression. Investigation of the observed subgroups in the treatment arm and observed data likelihood-ratio tests on nested subgroup models provided evidence of an interaction between node-status and presence of ulceration. The model is $\log(T_i \mid G_i = g) = \beta_{0g} + \psi_g R_i + \beta_1 \text{Breslow}_i + \beta_{2g} I\,[\text{Ulceration}]_i + \beta_3 I\,[\text{Trunk}]_i + \sigma\varepsilon$, $g = 0,1$, where $\psi_0 = 0$. The covariates have improved precision for the test of the null hypothesis H_0: $\psi_1 = 0$ (Table III(d): $\hat{\psi}_1 = 1.135$, p-value = 0.0024). Increased Breslow thickness and location of the melanoma on the trunk negatively impact DDFS in patients with and without nodal metastases. Interestingly, presence of ulceration portends poor DDFS in node-negative patients but not in node-positive ones, and this phenomenon is significant (p-value for test of $\beta_{20} = \beta_{21}$ is 0.024). To our knowledge, this interaction has not been previously reported. When covariates are added to the corresponding ITT analysis of effectiveness (which cannot accommodate subgroup by covariate interactions), there is no significant benefit of the new standard of care to either DFS or DDFS likely because of the small proportion of node-positive patients.

6. Discussion

We have developed methodology for latent subgroup analysis of a right-censored survival endpoint via a parametric AFT model that builds on prior work in all-or-none treatment noncompliance. An advantage of our approach is that the computational procedure is transparent and readily implemented in most statistical softwares without additional programming steps. It incorporates various levels of covariate effects and leads to maximum likelihood estimates, qualities that are desirable in most clinical applications. We have framed the method in the context of a two-subgroup scenario where a biological treatment effect is not defined in one group whose care is the same under both randomization assignments. Extensions to multiple untreatable subgroups are straightforward, and relaxations of the exclusion restriction can be explored, for example, through sensitivity analysis [23].

Latent subgroup analysis is relevant to a wide array of clinical trials, particularly in oncology and biomarker research, where a pathology report or diagnostic test that distinguishes subgroups of patients is available only in one arm of the study. Aspects of our framework also pertain to current research in principal stratification and auxiliary variables methods for causal inference [24, 25] in which estimation of causal treatment effects requires adjustment for information obtained post-randomization.

Our research direction was influenced in part by prior analysis of MSLT-I that supported the use of a parametric model, but many applications will fail to satisfy this assumption. We envision a semiparametric extension that would allow arbitrary residual error distributions in each subgroup. A second area of future research emanates from the fact that a large proportion of deaths (25 per cent) in MSLT-I is attributable to causes outside melanoma. The presence of competing risks may violate our assumption of noninformative censoring, and the current model could be extended to accommodate this scenario.

Acknowledgments

We thank Dr. Donald L. Morton of the John Wayne Cancer Institute at Saint John's Health Center in Santa Monica, CA and the Multicenter Selective Lymphadenectomy Trials Group (NCI grant CA-29605) for permission to use their data and for their helpful comments and critiques of this project. We also thank Dr. He-Jing Wang of UCLA for her help with data preparation and the associate editor and reviewers whose helpful comments lead to improvements in the paper. Gang Li's research is supported in part by NIH grants CA-016042 and P01-AT003960.

References

1. Morton DL, Thompson JF, Essner R, Elashoff R, Stern SL, Nieweg OE, Roses DF, Karakousis CP, Mozzillo N, Reingten D, Wang H, Glass EC, Cochran AJ. The Multicenter Selective Lymphadenectomy Trial Group. Validation of the accuracy of intraoperative lymphatic mapping and sentinel lymphadenectomy for early-stage melanoma: a multicenter trial. Annals of Surgery. 1999; 230(4):453–463. [PubMed: 10522715]

2. Morton DL, Cochran AJ, Thompson JF, Elashoff R, Essner R, Glass EC, Mozzillo N, Nieweg OE, Roses DF, Hoekstra HJ, Karakousis CP, Reintgen DS, Coventry BJ, Wang H. The Multicenter Selective Lymphadenectomy Trial Group. Sentinel node biopsy for early-stage melanoma: accuracy and morbidity in MSLT-I, an international multicenter trial. Annals of Surgery. 2005; 242(3):302–313. [PubMed: 16135917]

3. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. Journal of the American Statistical Association. 1996; 91(434):444–455.

4. Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. Annals of Statistics. 1997; 25(1):305–327.

5. Sommer A, Zeger SL. On estimating efficacy from clinical trials. Statistics in Medicine. 1991; 10(1):45–52. [PubMed: 2006355]

6. Palmgren, J.; Goetghebeur, E. Methods incorporating compliance in treatment evaluation. In: Geller, N., editor. Advances in Clinical Trial Biostatistics. CRC Press; Boca Raton: 2004.

7. Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. Statistical Methods in Medical Research. 2005; 14(4):369–395.10.1191/0962280205sm403oa [PubMed: 16178138]

8. Frangakis CE, Rubin DB. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. Biometrika. 1999; 86(2):365–379.10.1093/biomet/86.2.365.

9. Imbens GW, Rubin DB. Estimating outcome distributions for compliers in instrumental variables models. The Review of Economic Studies. 1997; 64(4):555–574.

10. Abadie A. Bootstrap tests for distributional treatment effects in instrumental variable models. Journal of the American Statistical Association. 2002; 97(457):284–292.

11. Loeys T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Biometrics. 2003; 59(1): 100–105.10.1111/1541-0420.00012 [PubMed: 12762446]

12. Cuzick J, Sasieni P, Myles J, Tyrer J. Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2007; 69(4):565–588.

13. Follmann D. On the effect of treatment among would-be treatment compliers: an analysis of the multiple risk factor intervention trial. Journal of the American Statistical Association. 2000; 95(452):1101–1109.

14. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics—Theory and Methods. 1991; 20(8): 2609–2631.

15. Mark SD, Robins JM. A method for the analysis of randomized trials with compliance information: an application to the multiple risk factor intervention trial. Controlled Clinical Trials. 1993; 14:79–79. [PubMed: 8500308]

16. Korhonen PA, Laird NM, Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC study. Statistics in Medicine. 1999; 18(21):2879–2897.10.1002/(SICI)1097-0258(19991115)18:21-2879::AID-SIM190-3.0.CO;2-K [PubMed: 10523748]

17. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. 1977; 39(1):1–38.

18. Little RJ, Yau LHY. Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin’s causal model. Psychological Methods. 1998; 3:147–159.

19. O’Malley AJ, Normand SLT. Likelihood methods for treatment noncompliance and subsequent nonresponse in randomized trials. Biometrics. 2005; 61(2):325–334.10.1111/j.1541-0420.2005.040313.x [PubMed: 16011678]

20. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2009. Available from: http://www.R-project.org

21. Cantor, A. SAS Survival Analysis Techniques for Medical Research. SAS Publishing; Cary, NC, U.S.A: 2003.

22. Morton DL, Thompson JF, Cochran AJ, Mozzillo N, Elashoff R, Essner R, Nieweg OE, Roses DF, Hoekstra HJ, Karakousis CP, Reintgen DS, Coventry BJ, Glass EC, Wang H. The Multicenter Selective Lymphadenectomy Trial Group. Sentinel-node biopsy or nodal observation in melanoma. New England Journal of Medicine. 2006; 355(13):1307–1317.10.1056/NEJMoa060992 [PubMed: 17005948]

23. Loeys T, Goetghebeur E. A sensitivity analysis for causal parameters in structural proportional hazards models. Statistics and Operations Research Transactions. 2003; 27(1):31–40.

24. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002; 58(1):21–29. [PubMed: 11890317]

25. Joffe MM, Small D, Hsu CY. Defining and estimating intervention effects for groups that will develop an auxiliary outcome. Statistical Science. 2007; 22(1):74–97.

Table I

Duality between the latent subgroup framework of MSLT-I and the all-or-none treatment noncompliance framework when controls have no access to active treatment.

	MSLT-I		Noncompliance	
Randomization	**Subgroup**	**Treatment receipt**	**Subgroup**	**Treatment receipt**
Treatment	Node-positive	Active treatment	Complier	Active treatment
Treatment	Node-negative	Control treatment	Noncomplier	Control treatment
Control	?	Control treatment	?	Control treatment

In MSLT-I, active treatment represents immediate lymphadenectomy, and the control treatment represents observational follow-up.

Table II

Large sample (a) and small sample (b) performance of the method in estimating the biological efficacy (ψ_1) in a Weibull model under various censoring rates (CR) and treatable subgroup proportions (p), (c) simulation of MSLT-I conditions (CR>75 per cent, p = 0.2) under three different efficacy scenarios.

CR	True p	$\hat{\psi}_1$	Coverage	Simulation variance	Estimated variance
(*a*) *Large sample simulation* (N = 1000, *true* ψ_1 = 0.85)					
40 per cent	0.3	0.8432	94.7	0.0420	0.0421
	0.5	0.8551	95.1	0.0197	0.0199
	0.8	0.8502	94.7	0.0106	0.0106
60 per cent	0.3	0.8539	95.3	0.0477	0.0487
	0.5	0.8521	95.3	0.0243	0.0247
	0.8	0.8492	95.5	0.0141	0.0144
(*b*) *Small sample simulation* (N = 400, *true* ψ_1 = 0.85)					
40 per cent	0.3	0.8379	93.9	0.1176	0.1108
	0.5	0.8504	93.8	0.0552	0.0500
	0.8	0.8523	95.9	0.0257	0.0266
60 per cent	0.3	0.8414	93.5	0.1325	0.1255
	0.5	0.8565	93.8	0.0632	0.0645
	0.8	0.8544	94.6	0.0364	0.0362
(*c*) *Simulation of MSLT-I* (N = 1300)					
True ψ_1					
1.00		0.9842	94.7	0.1023	0.1012
0.50		0.5012	95.6	0.0950	0.0957
0.00		0.0071	95.7	0.0923	0.0966

Table III

(a) Pertinent summary statistics on the survival endpoints of interest. Disease-free survival (DFS) is defined as time until a clinically detectable recurrence at any site (no melanoma deaths occurred without recurrence). Distant-disease-free survival (DDFS) is specific to recurrence at a distant site and includes death from melanoma without a distant recurrence. The two subgroups are patients with sentinel-node metastases (node +) and those absent metastases (node −). (b) Results from the two-sample ITT analysis estimating the effectiveness of the new standard of care; the parameter estimates are from Weibull accelerated failure time models. (c) Results from a model to estimate efficacy of lymphadenectomy on node-positive patients without covariates. (d) Follow-up analysis of DDFS to investigate covariate effects.

	Treatment, node + (*N* = 122)	**Treatment, node − (*N* = 642)**	**Control (*N* = 500)**
(*a*) *Summary statistics on survival endpoints* (*N* = 1264)			
Recurrence at any site—per cent	44.3	16.0	26.8
Distant recurrence—per cent	36.9	12.3	17.8
DFS (months)—median	37.9	59.1	54.5
DDFS (months)—median	43.8	59.8	59.9
(*b*) *Two-sample intention-to-treat test of effectiveness* (*N* = 1264)			
	Effectiveness	95 per cent interval	*p*-Value
DFS	0.344	(0.0782, 0.609)	0.0112
DDFS	0.0708	(−0.201, 0.343)	0.610
(*c*) *Two-sample test of efficacy in node-positive subgroup* (*N* = 1264)			
	Efficacy ($\hat{\psi}_1$)	95 per cent interval	*p*-Value
DFS	1.593	(1.167, 2.019)	2.4E–13
DDFS	0.937	(0.119, 1.755)	0.0247
(*d*) *Regression analysis of DDFS with covariates* (*N* = 1147)			
Shared parameters	Estimate	95 per cent interval	*p*-Value
Breslow thickness	−0.528	(−0.739, −0.318)	8.6E–7
Trunk (site)	−0.486	(−0.766, −0.205)	6.8E–4
Node +			
Treatment efficacy	1.135	(0.402, 1.869)	0.0024
Ulceration	0.179	(−0.377, 0.735)	0.527
Node −			
Ulceration	−0.639	(−1.007, −0.270)	6.8E–4