We introduce a simple multiplicative SNMM [6–8]

$$\log E[T_i(\overline{S}_i(t),0)|H_i(t),S_i(t)]-\log E[T_i(\overline{S}_i(t-1),0)|H_i(t),S_i(t)]=\beta_0 S_i(t) \tag{1}$$

where β_0 is the constant (across t) incremental causal effect of a final treatment $S_i(t)$ at time t on the potential outcome $T_i(\overline{S}_i(t-1),0)$ following a patient's actual treatment through times $0,\ldots,t-1$ and control treatment after $t-1$. Under this constant treatment effect model (1), β_0 multiplied by M (number of visit time) can be interpreted as the average causal treatment effect that would be realized if all patients had continued to comply with the treatment to which they were assigned. Robins [6–8] proposed the estimation method of β_0, the so-called, g-estimation method, under the assumption of sequential conditional independence for any t and k with $k\leqslant t-1$

$$T_i(\overline{S}_i(k),0)\perp\!\!\!\perp S_i(t)|H_i(t) \tag{2}$$

which states that, when $k\leqslant t-1$, treatment $S_i(t)$ is independent of the potential outcomes $T_i(\overline{S}_i(k),0)$, given the observed history up to time t, $H_i(t)$. In practice, we would not expect this assumption to be precisely true, but given a rich collection of prognostic factors that influence a patient's decision to comply at time t recorded in $H_i(t)$, it may well be approximately true. Robins [6–8] has referred to (2) as the assumption of no unmeasured confounders.

3.2. *Estimation of β_0 via the intensity score method*

Brumback *et al.* [9] proved that the SNMM treatment effect, that is, g-estimator of β_0, can be obtained by the intensity score method, in which outcomes are regressed on the cumulative intensity score. We utilize their results and consider the accelerated failure time (AFT) model to obtain the consistent estimator of β_0 in the multiplicative SNMM (1). Here, we assume that the observed event time T_i is subject to independent random censoring such as an end-of-study censoring, where T_i for censored subjects is either the time until dropout or the time until end of study.

We assume the following exponential regression model (log-linear model) for T_i [20]:

$$\log T_i=\mu+\beta_I\sum_{t=0}^{M-1}\hat{I}_i(t)+\varepsilon_i \tag{3}$$

where μ is the intercept parameter, $I_i(t)=S_i(t)-E[S_i(t)|H_i(t)]$ is the intensity score at time t, and ε_i follows the extreme value distribution. For binary treatment $S_i(t)$, the time-dependent intensity score $I_i(t)$ measures departures of actual treatment from the propensity score $\Pr[S_i(t)|H_i(t)]$ of Rosenbaum and Rubin [21]. Since the propensity score is usually unknown, it must be estimated from the data. If we assume a parametric model for $\Pr[S_i(t)|H_i(t)]$ such as

$$\operatorname{logit}\Pr[S_i(t)=1|H_i(t)]=\theta^{\mathrm{T}}H_i(t) \tag{4}$$

then the intensity score at time t can be estimated by $\hat{I}_i(t)=S_i(t)-E[S_i(t)|H_i(t);\hat{\theta}]$, where $\hat{\theta}$ is the maximum likelihood estimator of θ under model (4). Here we assume that the intensity score at time t is not equal to zero with probability 1 for each patient, that is, $\hat{I}_i(t)\neq 0$ for any t. This assumption will be satisfied unless there is a covariate level $H_i(t)$ such that all patients with that level of the covariate are certain to receive the treatment.

The estimate for β_I in model (3) can be obtained *via* the ordinary-weighted least-squares (WLS) method. However, the cumulative intensity score is generally uncorrelated with the cumulative propensity score, although $E[I_i(t)E(S_i(t)|H_i(t))]=0$ for any t. Therefore, as Brumback *et al.* [9]

DOI: 10.1002/sim

have pointed out in the case of linear model, to obtain a consistent estimator of β_0 *via* the WLS method for model (3), the correction term $N\beta_I C$ must be subtracted from the WLS estimating function, where $C=(1/N)(\sum_{t=0}^{M-1}\hat{I}_i(t))\omega_i(\sum_{t=0}^{M-1}E(S_i(t)|H_i(t);\hat{\theta}))$ and $\omega_i=\exp(-\mu)\cdot T_i\cdot\exp(-\beta_I\sum\hat{I}_i(t))$. The corrected estimating function for model (3) is

$$U(\mu,\beta_I)\equiv\sum_i^N(-d_i\sum\hat{I}_i(t)+\sum\hat{I}_i(t)\exp[\log t_i-\mu-\beta_I\sum\hat{I}_i(t)])-N\beta_I C=0 \tag{5}$$

where d_i is the event indicator that takes the value of one if the subject failed and zero if the subject is censored. In Appendix A, we show the proof that the correction term must be subtracted from the WLS estimating function to obtain a consistent estimator.

Our estimating function has the form $\sum_i^N U_i(\gamma)=0$, where $\gamma=(\mu,\beta_I,\theta)$ represents the intercept, coefficient of the intensity score, and parameter used to model the propensity score. To correct for having the estimated θ, the asymptotic variance of $\hat{\gamma}$ was obtained by using a sandwich estimator, which was computed as

$$[\hat{E}(\partial U_i/\partial\gamma)]^{-1}[\hat{E}(U_iU_i^{\mathrm{T}})\hat{E}(\partial U_i/\partial\gamma)]^{-1,\mathrm{T}}/N$$

where the estimated expectations were computed using the empirical distribution of the sample.

3.3. Time-dependent treatment effects

An extension of the multiplicative SNMM (1) is to allow the treatment effects to vary across time,

$$\log E[T_i(\overline{S}_i(t),0)|H_i(t),S_i(t)]-\log E[T_i(\overline{S}_i(t-1),0)|H_i(t),S_i(t)]=\beta_0(t)S_i(t) \tag{6}$$

where $\beta_0(t)$ is the causal parameter at each time t. Since $\beta_0(t)$ is the incremental causal effect of a final treatment $S_i(t)$ at time t, the cumulative effect $\sum_{k=0}^{t}\beta_0(k)$ is the average causal treatment effect that would be realized if all patients had continued to comply with the treatment to which they were assigned until time t. Assuming the consistency assumption and the sequential conditional independence (2), the time-dependent causal parameters in model (6) are consistently estimated by fitting the following model [9]

$$\log T_i=\mu+\sum_{t=0}^{M-1}\beta_I(t)\hat{I}_i(t)+\varepsilon_i \tag{7}$$

where the correction term $\sum_i[\hat{I}_i(t)\cdot\omega_i\cdot\sum_{t=0}^{M-1}\{\beta_I(t)E(S_i(t)|H_i(t);\hat{\theta})\}]$ must be subtracted from the WLS estimating function for model (7).

4. SIMULATION STUDY

4.1. Simulation design

To evaluate the performance of the AT, ITT, g-estimation (see Section 4.2) and intensity score methods, we carried out simulation studies under non-random non-compliance. We simulated data from two treatment groups, coded as $R=0$ (control treatment) or $R=1$ (test treatment). About equal sample size of 1000 for each group was randomly generated (total sample size was 2000).

DOI: 10.1002/sim

The simulations were based on 1000 replications so that the estimated coverage probability of a true 95 per cent CI would have a simulation accuracy of approximately 1.35 per cent.

For each subject i ($i=1,\ldots,2000$), a baseline covariate L_i was generated from the normal distribution with mean of 2 and variance of 1. The potential baseline failure time U_i was generated from the following exponential model:

$$U_i = U_0 \exp(\alpha_0 + \alpha_1 L_i) \tag{8}$$

where U_0 was an exponential random number with mean of 1 and $(\alpha_0, \alpha_1) = (3.2, -0.5)$ so that the larger the value of L_i, the shorter the baseline failure time U_i. We evaluated the treatment actually received $S_i(t)$ at three time points $t=0$, 2, and 4, where all subjects were assumed to take the assigned treatment at $t=0$ ($S_i(0)=R_i$) and the treatment crossover occurred at $t=2$ and 4 according to the following model:

$$\text{logit Pr}[S_i(t)] = \gamma_0 + \gamma_1 L_i + \gamma_2 S_i(t-2) \tag{9}$$

where $(\gamma_1, \gamma_2) = (1.2, 4.5)$ so that patients with poor prognosis and taking the test treatment at previous time point tended to receive the test treatment. The non-compliance rate was adjusted by the value of the intercept parameter γ_0, where two settings were considered: 45 per cent ($R=0$) *versus* 15 per cent ($R=1$) and 30 per cent ($R=0$) *versus* 10 per cent ($R=1$). In this non-compliance rate, the subject was considered as a non-complier when the subject received another treatment even once during the study period.

The observed failure time T_i was calculated from the SAFT model

$$U_i = \int_0^{T_i} \exp[-\psi_0 S_i(t)]\,dt \tag{10}$$

where ψ_0 is the causal treatment effect, which was set to $\psi_0 = 0.5$. The observed failure time T_i was censored at the fixed censoring time C, where $C=5$, 70, and ∞, so that the overall censoring proportion was nearly 90, 30, and 0 per cent, respectively.

Simulations were evaluated in terms of the bias, mean-squared error (MSE), mean length of the 95 per cent CI (length), 95 per cent coverage probability (CP), power for rejecting the null hypothesis, and α-error.

4.2. g-estimation (semi-parametric randomization-based analysis)

A semi-parametric randomization-based approach to estimate the causal effect has been developed by Robins and coworkers [10, 12]. For time-to-event outcomes, their approach is based on the causal AFT model (10), which relates a patient's observed event time T_i to the potential baseline event time U_i, that would have been observed if no treatment had been given, and the treatment actually received $S_i(t)$ *via* a causal parameter ψ_0. Note that if $S_i(t) \equiv 0$, then equation (10) gives $T_i = U_i$ as expected, while if $S_i(t) \equiv 1$, (10) gives $T_i = U_i \exp(\psi_0)$. Therefore, equation (10) implies that for continuous treatment the potential event time U_i is prolonged by the factor $\exp(\psi_0)$. A positive value of ψ_0 represents a beneficial treatment effect.

To estimate the causal parameter ψ_0, they choose to avoid all assumptions about both observed and unobserved factors that influence an individual's decision to comply such as (2), while comparing outcomes based only on the treatment groups randomized by design, that is, their analyses are randomization-based analysis. The key to understanding their estimation method

DOI: 10.1002/sim

(g-estimation) is to realize that U_i is a baseline variable identically distributed across the randomized groups. We define $U_i(\psi)$ to equal the right-hand side of (10) for given ψ. We also define $Z(\psi)$ to be a test statistic comparing the distribution of $U(\psi)$ in the two randomized groups, where we will use the log-rank test. The point estimate of ψ_0 is the value for which $Z(\psi)=0$, and this can be found by a search over a grid. A $100(1-\alpha)$ per cent confidence interval for ψ_0 is the range of values for which $|Z(\psi)|<z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution. One attractive point of this approach is that at the null value, it is non-parametric, because $U_i(0)=T_i$; hence, $Z(0)$ is the usual ITT log-rank test statistic.

However, if T_i is a censored time, then $U_i(\psi)$ is censored at

$$D_i(\psi)=\int_0^{C_i}\exp[-\psi S_i(t)]\,dt$$

where C_i is defined as the time between subject i's randomization and the fixed end of the follow-up date. Although C_i is known for uncensored as well as censored subjects, $D_i(\psi)$ is a function of $S_i(t)$ and may depend on the underlying prognosis. Therefore, even when censoring on the T-scale is non-informative, that is, an administrative censoring, censoring on the U-scale is likely to be informative, if $\psi_0\neq 0$ and there is non-random non-compliance. Thus, we cannot replace T_i by $X_i=\min(T_i, C_i)$ to calculate the pseudo-treatment-free event time.

To avoid this problem, Robins and Tsiatis [10] defined a new censoring time $C_i(\psi)=C_i$ if $\psi\leqslant 0$ and $C_i(\psi)=C_i\exp(-\psi)$ if $\psi>0$, according to the direction of treatment effect. For given ψ, let $X_i(\psi)=\min[C_i(\psi), U_i(\psi)]$ and $\Delta_i(\psi)=I[U_i(\psi)>C_i(\psi)]$ to be the new follow-up time and censoring indicator, respectively. $X_i(\psi)$ is observable since $T_i\geqslant C_i$ implies $U_i(\psi)>C_i(\psi)$. Because any function of $\{U_i(\psi), C_i\}$ is independent of random treatment assignment R_i, we have $\{U_i(\psi_0), \Delta_i(\psi_0)\}\coprod R_i$, where the symbol $\coprod$ means independent.

4.3. *Results of simulations*

Table II shows the results. The constant treatment effect model (3) with $M=3$ was applied in the intensity score analysis, where a logistic regression model (9) was used for the estimation of the propensity score at $t=2$ and 4. From Table II we see that both the AT and ITT estimates were largely biased toward the null in all situations (true value of treatment effect$=0.5$). The biases increased as the difference of non-compliance proportion between groups increased and as the censoring proportion decreased. The α-errors for the ITT estimate were close to the nominal level of 5 per cent, reflecting that the ITT approach provides a valid test for the null hypothesis of no treatment effect even in the presence of non-random non-compliance.

As expected, the g-estimates performed well in all situations, because the data generation process was based on the SAFT model (10). Note that the powers for the ITT and g-estimate were about the same, reflecting that even though the g-estimation approach uses non-compliance information it does not increase the power against the null hypothesis when compared with the ITT approach.

The intensity score estimates were nearly unbiased and their coverage probabilities were close to the nominal level of 95 per cent in all situations. The α-errors were controlled under the correctly specified parametric model (9). The intensity score estimates gave smaller MSE and narrower confidence intervals than those of the g-estimates, except in the censoring proportion$=90$ per cent. The powers were slightly increased compared with the g-estimates.

DOI: 10.1002/sim

Table II. Results of simulation studies for AT, ITT, g-estimation, and IS method.

Method	Non-compliance	Censoring (per cent)	Bias	MSE	CI length	95 per cent CP	Power	α Error
AT	45 *versus* 15 per cent	0	−0.415	0.175	0.183	0.0	44.3	100.0
		30	−0.405	0.170	0.216	0.0	41.4	100.0
		90	−0.230	0.074	0.564	64.0	47.5	30.4
	30 *versus* 10 per cent	0	−0.363	0.134	0.180	0.0	84.3	95.0
		30	−0.353	0.127	0.212	13.7	77.7	100.0
		90	−0.190	0.056	0.559	73.7	57.8	22.2
ITT	45 *versus* 15 per cent	0	−0.305	0.095	0.176	0.0	99.4	5.6
		30	−0.295	0.090	0.209	0.1	97.0	5.8
		90	−0.138	0.041	0.567	82.6	71.1	4.8
	30 *versus* 10 per cent	0	−0.253	0.066	0.177	0.0	99.9	4.3
		30	−0.243	0.062	0.209	0.4	99.8	6.1
		90	−0.108	0.033	0.566	87.2	78.4	4.6
g-estimation	45 *versus* 15 per cent	0	0.006	0.013	0.541	95.1	99.4	5.6
		30	0.008	0.018	0.615	94.3	97.0	5.8
		90	0.001	0.037	0.854	94.6	70.9	4.9
	30 *versus* 10 per cent	0	0.001	0.009	0.454	95.4	99.9	4.3
		30	0.003	0.015	0.509	97.0	100.0	6.1
		90	−0.001	0.032	0.825	96.5	78.3	4.6
Intensity score	45 *versus* 15 per cent	0	−0.046	0.005	0.274	95.7	100.0	5.2
		30	−0.019	0.005	0.287	94.5	100.0	4.6
		90	0.060	0.061	0.959	97.8	74.3	4.7
	30 *versus* 10 per cent	0	−0.045	0.004	0.257	95.2	100.0	4.5
		30	−0.018	0.005	0.262	95.2	100.0	4.6
		90	0.053	0.046	0.840	98.0	82.4	4.7

AT, as-treated; MSE, mean-squared error; CI, confidence interval; CP, coverage probability.

5. ANALYSIS OF MEGA STUDY DATA

In the analysis of the MEGA study data, we divided the follow-up period into 10 time intervals with equal space (1 year). Patients were classified as a non-complier in a time interval if he/she switched to the other trial treatment at least once during the interval.

5.1. Estimation of the propensity score

To estimate the propensity score at each time t ($t=0,\ldots,9$), the logistic regression model (4) was used, in which four time-dependent factors as well as 12 baseline factors shown in Table I were included as covariates $H_i(t)$. For the four time-dependent factors, the most recent recorded values were included as covariates $H_i(t)$ in model (4). All TC values were excluded accounting for the multicollinearity of covariates. Among baseline factors, missing data were observed in the values of BMI (0.24 per cent), current smoking (0.18 per cent), and drinking (0.17 per cent). The missing values of BMI were imputed by the mean value of 23.8 (kg/m^2). The latter two factors were imputed by zero (no smoking and no drinking, respectively). Four time-dependent factors were three lipids (TG, HDL-C, and LDL-C) and treatment actually received before time t. For the

DOI: 10.1002/sim

Table III. Predictors of receiving the pravastatin treatment at $t=3$.

Predictors	Odds ratio	95 per cent CI
Baseline covariates		
Assigned treatment	4.645	3.536, 6.102
Age (years)	1.008	0.991, 1.026
Women	0.916	0.663, 1.264
BMI (kg/m^2)	1.008	0.968, 1.050
Current smoker	1.262	0.884, 1.800
Current drinking	0.932	0.684, 1.271
Medication history	1.484	1.086, 2.029
Hypertension	1.169	0.915, 1.493
Diabetes	1.247	0.938, 1.658
TG (mg/dL)	1.001	0.998, 1.003
HDL-C (mg/dL)	0.992	0.974, 1.010
LDL-C (mg/dL)	1.013	1.003, 1.023
Time-dependent covariates		
TG (mg/dL) at $t=2$	1.003	1.001, 1.005
HDL-C (mg/dL) at $t=2$	1.030	1.014, 1.046
LDL-C (mg/dL) at $t=2$	1.010	1.001, 1.015
Treatment received at $t=2$	240.2	179.2, 321.7

CI, confidence interval; medication history: hypercholesterolemia medication history.

missing data of lipid values (21.5 per cent), the regression imputations were separately conducted, where 11 baseline factors, allocation group, and the last observed lipid value were included as covariates in each prediction model.

Table III shows the odds ratio of each factor associated with receiving the pravastatin treatment at time $t=3$. The results for other time points (not shown) were essentially similar to those shown in Table III. For the baseline covariates, patients who were assigned to the pravastatin group and have hypercholesterolemia medication history tended to receive the pravastatin treatment. As expected, the previous use of pravastatin also predicted the use of pravastatin subsequently.

5.2. *Estimation of treatment effect adjusting for treatment changes*

Table IV shows the estimates of treatment effect by several methods. Hazard ratios for stroke event, which was one of the secondary endpoints in the MEGA study, were also presented. Analysis models for stroke were the same as those for CHD events, and similar results for factors associated with receiving the pravastatin treatment were observed (not shown) as shown in Table III. For both CHD and stroke events, two analyses were conducted, where each endpoint was evaluated at 5 or 10 years, respectively. Two intensity score estimates were obtained: one (intensity score 1) was the constant treatment effect by applying model (3) and the other (intensity score 2) was the cumulative treatment effect by applying model (7).

Both the intensity score and g-estimation methods gave the larger treatment effects for pravastatin than the ITT ones for all endpoints. The adjustment effects were larger in the stroke events. The statistically significant effect in the stroke event at 10 years was observed by the intensity score 1 (hazard ratio=0.51; 95 per cent CI: 0.28–0.95). The results from intensity score 2, in particular

DOI: 10.1002/sim

Table IV. Estimates of treatment effect for CHD and stroke events.

	CHD				Stroke			
	5 years		10 years		5 years		10 years	
Method	HR	95 per cent CI	HR	95 per cent CI	HR	95 per cent CI	HR	95 per cent CI
ITT	0.70	0.50, 0.97	0.67	0.49, 0.91	0.65	0.43, 0.97	0.83	0.57, 1.21
Intensity score 1	0.68	0.44, 1.05	0.59	0.36, 0.99	0.44	0.25, 0.79	0.51	0.28, 0.95
Intensity score 2	0.68	0.46, 1.02	0.66	0.27, 1.60	0.53	0.31, 0.90	0.45	0.17, 1.21
g-estimation	0.65	0.30, 0.91	0.64	0.39, 0.83	0.54	0.26, 0.87	0.63	0.33, 1.26

HR, hazard ratio; CI, confidence interval; intensity score 1, constant treatment effect from model (3); intensity score 2, cumulative treatment effect from model (7); g-estimation, semi-parametric randomization-based analysis using model (10).

at 10 years, gave the wider confidence intervals than those from intensity score 1, which probably reflects the sparse data problems in estimating $\beta_I(t)$. The confidence intervals for the g-estimates contained the null value of 1 whenever the ITT result was not significant.

6. DISCUSSION

In this paper, we developed the intensity score approach for time-to-event outcomes with censoring to estimate the causal treatment effect in the presence of non-random non-compliance. The proposed approach has three major advantages over the g-estimation based on the SAFT model (10). The first advantage is that an artificial recensoring scheme (Section 4.2) is necessary requirement for the g-estimation to account for administrative censoring correctly, while the proposed approach can treat the censoring uniquely within the framework of standard regression models. The rationale for recensoring in the g-estimation is that if the potential baseline failure time U_i is independent of treatment assignment, the same should be true for any function of U_i and C_i since C_i is a baseline covariate. Therefore, there are several choices for an observable random variable that is a function of $\{U_i, C_i\}$ as a basis for inference [13, 16, 22].

The second major advantage of the proposed approaches is that they can be easily extended to the estimation of time-dependent treatment effects such as (6), where the technique of g-estimation has been difficult to apply in practice to the multi-parameter model. Although the constant treatment effect model (1) is very simple, model (6) is more robust to the estimation of dynamic sequential treatments conditional on past medical history. This robustness property of model (6) will be compromised with the sacrifice of the precision as shown in Table IV. To avoid the sparse-data problems, Brumback *et al.* [9] proposed the use of parametric constraints among the $\beta_I(t)$ such as $\beta_I(t)=a_0+a_1t$ depending on context.

The third advantage is its ease of application, that is, the g-estimate can be obtained in three steps: we compute propensity scores, derive intensity scores, and fit an ordinary regression model for any outcome variable, although the correction term must be subtracted from the estimating function to obtain the consistent estimator.

Nevertheless, the g-estimation has a number of advantages over the proposed approach. First one is that it is a semi-parametric randomization-based approach, that is, it preserves the validity of tests of the null hypothesis regardless of what determinants of outcome have influenced a

DOI: 10.1002/sim

patient's decision to comply. Furthermore, the g-estimation provides estimated effects that are of the same sign as the ITT effect and that are only statistically significant if the ITT analysis is statistically significant. In relation to this point, a major drawback of the intensity score approach is that one must be able to specify a correct model for the conditional probability of treatment, $\Pr[S_i(t)|H_i(t)]$, for each t up to the end of follow-up, although the increase of power will be anticipated. Unfortunately, the assumption of no unmeasured confounders (2) is a non-identifiable assumption and is not testable from the observed data. Furthermore, even when assumption (2) is approximately true, we require strong modeling assumptions, since there are many covariates in $H_i(t)$. It is unlikely that these modeling assumptions would be precisely correct. In the MEGA study, many clinically important prognostic factors were measured and all of them were used as covariates to estimate the propensity score at each time. In addition to the prediction model shown in Table III, the analyses based on other prediction models, such as a parsimonious model using a variable selection procedure or full models in which time-dependent covariates, were entered as the difference from the baseline or the absolute past two values, and the intensity score estimates were shown to be insensitive to the selection of the prediction models conditional on the measured covariates.

Another advantage of the g-estimation over the intensity score approach is that one can use the SAFT model (10) to estimate the effect of a treatment on outcome in studies, where at each time t there is a covariate level such that all patients with that level of the covariate are certain to receive the identical treatment. For example, this circumstance implies that the intensity score approach should not be used for the analysis of non-compliance data, in which treatment switching was observed in only one group, because the intensity score at each time will be zero for patients in the complete compliance group. Robins [23] and Robins *et al.* [24] discussed a similar problem, that is, structural zero, for the adjustment of time-dependent confounding and showed that the IPTW (inverse probability of treatment weighted) estimators, which are based on the propensity score, are biased for the data with structural zero.

As Brumback *et al.* [9] have discussed, the intensity score approach resembles the IPTW estimation method based on the marginal structural model (MSM). Although the MSM is useful for estimating the causal effect of the pre-specified treatment regime such as always treat or treat on alternate month [23, 24], it is much less useful for modeling the interaction of treatment with a time-dependent covariate and for estimating the effect of a dynamic treatment plan in which the treatment on a visit depends on a subject's evolving covariate history. It is important to recognize that actual medical treatment regimes including non-compliance data are usually dynamic, and the SNMM is more suitable for parametrizing such dynamic effects. Another difference between the SNMM and the MSM is that the latter makes fewer assumptions than the former by not requiring treatment effects to be constant across strata of covariate history, because the IPTW estimators can be interpreted as standardized parameters [24, 25]. Thus, in theory, the IPTW estimator is more robust than the intensity score one.

In the analyses of the MEGA study data, we observed the larger adjustment effects in the stroke events in spite of the fact that factors associated with non-compliance were nearly the same for CHD and stoke events in each group. The explanatory analyses among the non-compliant cases were conducted to investigate the relation between the non-compliance rate (/year) of each case and the occurrence of each event. These analyses showed that, in the diet group, the effect of non-compliance rate on the non-occurrence of stroke events (odds ratio = 144; 95 per cent CI: 1.3–∞; 5 stroke events among 865 non-compliant cases) was larger than that of CHD events (odds ratio = 14.5; 95 per cent CI: 1.7–150; 19 CHD events among 844 non-compliant cases), while,

DOI: 10.1002/sim

in the pravastatin group, the effect of non-compliance rate on the occurrence of stroke events (odds ratio = 5.7; 95 per cent CI: 1.2–26; 16 stroke events among 2440 non-compliant cases) was also larger than that of CHD events (odds ratio = 1.3; 95 per cent CI: 0.3–5.3; 20 CHD events among 2441 non-compliant cases). These facts may explain the larger discrepancy between the ITT estimate and the causal one observed in stroke events.

In the MEGA study, like any other clinical trial, dropout of patients during the study period was observed. In addition to the usual loss to follow-up cases, there was another problem of dropouts due to the refusal of further follow-up at 5 years [17, 18]. In this paper, we considered all these dropout cases as non-informative censoring cases. Because observed dropout proportions were not different among treatment groups (loss to follow-up: 546/3966 = 0.14 in diet group and 594/3866 = 0.15 in pravastatin group; refusal of follow-up by patients: 278/3966 = 0.07 in diet group and 270/3866 = 0.07 in pravastatin group), the effect of these dropouts on the comparison of treatment group may seem to be small. However, these non-administrative censorings may be informative and hence a source of selection bias. To adjust for selection bias due to non-administrative censoring, the IPCW (inverse probability censoring weighted) method has been proposed [26–28]. The underlying idea of the IPCW method is to base estimation on the observed outcomes but weight them to account for the probability of being uncensored. We analyzed the MEGA study data using the IPCW method which can adjust for some types of dependent censorings, and confirmed that there were no large differences between the ITT estimates and the IPCW ones for both CHD and stroke events [29]. Our intensity score method can also incorporate the IPCW method, and this will be a future work.

APPENDIX A

We show that the correction term $N\beta_I C$ must be subtracted from the WLS estimating function to obtain a consistent estimator of β_0 in (1), where $C=(1/N)(\sum_{t=0}^{M-1}\hat{I}_i(t))\omega_i(\sum_{t=0}^{M-1}E(S_i(t)|H_i(t);\hat{\theta}))$ and $\omega_i=\exp(-\mu)\cdot T_i\cdot\exp(-\beta_I\sum\hat{I}_i(t))$. We define the 'estimated' potential outcome under no treatment:

$$\log\hat{T}_{0i}\equiv\log T_i-\beta_0\sum_{t=0}^{M-1}S_i(t)$$

Under model (1) and assumption (2), the estimated potential outcome is mean independent of future treatment given past history, which implies that $E[\hat{I}_i(t)\cdot w_i\cdot(\log\hat{T}_{0i}-\mu)]=0$, $t\leqslant M-1$. Therefore,

$$E\left(\sum_{t=0}^{M-1}\hat{I}_i(t)\cdot w_i\cdot\left[\log T_i-\mu-\beta_0\sum_{t=0}^{M-1}S_i(t)\right]\right)=0 \tag{A1}$$

Now, the WLS estimating equation that $\hat{\beta}_I$ solves under model (3) has unconditional mean zero if and only if $E(\sum_{t=0}^{M-1}\hat{I}_i(t)\cdot w_i\cdot[\log T_i-\mu-\beta_I\sum_{t=0}^{M-1}\hat{I}_i(t)])=0$. Substituting $\hat{I}_i(t)=S_i(t)-E[S_i(t)|H_i(t);\hat{\theta}]$ yields

$$E\left(\sum_{t=0}^{M-1}\hat{I}_i(t)\cdot w_i\cdot\left[\log T_i-\mu-\beta_I\sum_{t=0}^{M-1}S_i(t)+\beta_I\sum_{t=0}^{M-1}E[S_i(t)|H_i(t);\hat{\theta}]\right]\right)=0 \tag{A2}$$

DOI: 10.1002/sim

Comparing (A1) with (A2), it follows that $\hat{\beta}_I$ is consistent for β_0 if for any t, $\hat{I}_i(t) \neq 0$ and $E(\sum_{t=0}^{M-1} \hat{I}_i(t) \cdot w_i \cdot \sum_{t=0}^{M-1} E[S_i(t)|H_i(t); \hat{\theta}]) = 0$.

ACKNOWLEDGEMENTS

MEGA Study is supported by Sankyo Co. Ltd. We also thank reviewers for their comments, which led to a much improved version of the paper.

This research was supported in part by Grant-in-Aid for Scientific Research (A) No. 16200022. Research funds for MEGA Study were provided by the Japanese Ministry of Health, Labor and Welfare for the first 2 years of the study, and thereafter the study was funded by Sankyo Co Ltd, Tokyo.

REFERENCES

1. Efron B, Feldman D. Compliance as an explanatory variable in clinical trials (with Discussion). *Journal of the American Statistical Association* 1991; **86**:9–26.
2. Sommer A, Zegar SL. On estimating efficacy from clinical trials. *Statistics in Medicine* 1991; **10**:45–52.
3. Angrist JD, Imbens GW, Rubin DB. Identification of casual effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**:444–455.
4. Fisher LD, Dixon DO, Herson J, Frankowski RK, Hearron MS, Pease KE. Intention-to-treat in clinical trials. *Statistical Issues in Drug Research and Development*. Marcel Dekker, Inc.: New York, 1990; 331–350.
5. Koch GG, Edwards S. Clinical efficacy trials with categorical data. In *Biopharmaceutical Statistics for Drug Development*, Peace K (ed.). Marcel Dekker, Inc.: New York, 1988; 403–457.
6. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics* 1994; **23**:2379–2412.
7. Robins JM. Causal inference from complex longitudinal data. In *Latent Modelling with Applications to Causality*, Berkane M (ed.). Springer: New York, 1997; 69–117.
8. Robins JM. Correcting for non-compliance in equivalence trials. *Statistics in Medicine* 1998; **17**:269–302.
9. Brumback B, Greenland S, Redman M, Kiviat N, Diehr P. The intensity score approach for adjusting for confounding. *Biometrics* 2003; **59**:274–285.
10. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics* 1991; **20**:2609–2631.
11. Robins JM, Blevins D, Ritter G, Wolfsohn M. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology* 1992; **3**:319–336.
12. 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**:79–97.
13. Keiding N, Filiberti M, Esbjerg S, Robins JM, Jacobsen N. The graft versus leukemia effect after bone marrow transplantation: a case study using structural nested failure time models. *Biometrics* 1999; **55**:23–28.
14. White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: examples from the concorde trial. *Statistics in Medicine* 1999; **18**:2617–2634.
15. Korhonen PA, Laird NM, Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC study. *Statistics in Medicine* 1999; **18**:2879–2987.
16. Hernan MA, Cole SR, Margolick J, Cohen M, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and Drug Safety* 2005; **14**:477–491.
17. Management of Elevated Cholesterol in the Primary Prevention Group of Adult Japanese (MEGA) Study Group. Design and baseline characteristics of a study of primary prevention of coronary events with pravastatin among Japanese with mildly elevated cholesterol levels. *Circulation Journal* 2004; **68**(9):860–867.
18. Nakamura H, Arakawa K, Itakura H *et al.* Primary prevention of cardiovascular disease with pravastatin in Japan (MEGA Study): a prospective randomised controlled trial. *Lancet* 2006; **368**:1155–1163.
19. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics* 1978; **6**:34–58.
20. Cox DR, Oakes D. *Analysis of Survival Data*. Chapman & Hall: London, 1984.
21. Rosenbaum PR, Rubin DM. The central role of the propensity score in observational studies of casual effects. *Biometrika* 1983; **70**:41–55.

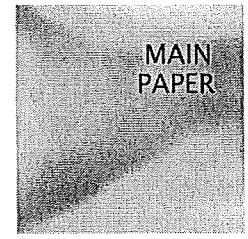
DOI: 10.1002/sim

22. Witteman JC, D'Agostino RB, Stijnen T *et al.* G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the Framingham heart study. *American Journal of Epidemiology* 1998; **148**:390–401.
23. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, Halloran ME, Berry D (eds). Springer: New York, 1999; 95–133.
24. Robins JM, Hernan MA, Brumback BA. Marginal structural models and casual inference in epidemiology. *Epidemiology* 2000; **11**:550–560.
25. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology* 2003; **14**:680–686.
26. Robins JM. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate makers. *American Statistical Association Proceedings of the Biopharmaceutical Section* 1993; 24–33.
27. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995; **90**:106–121.
28. Robins JM, Finkelstein DH. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with Inverse Probability of Censoring Weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**:779–788.
29. Yoshida M, Matsuyama Y, Ohashi Y for the MEGA Study Group. Estimation of treatment effect adjusting for dependent censoring using the IPCW method: an application to a large primary prevention study for coronary events (MEGA study). *Clinical Trials* 2007; **4**:318–328.

DOI: 10.1002/sim

PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. (in press)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/pst.290

Estimation of the marginal survival time in the presence of dependent competing risks using inverse probability of censoring weighted (IPCW) methods



Yutaka Matsuyama*,† and Takuhiro Yamaguchi
Department of Biostatistics, School of Health Sciences and Nursing, University of Tokyo, Bunkyo-ku, Tokyo, Japan

In medical studies, there is interest in inferring the marginal distribution of a survival time subject to competing risks. The Kyushu Lipid Intervention Study (KLIS) was a clinical study for hypercholesterolemia, where pravastatin treatment was compared with conventional treatment. The primary endpoint was time to events of coronary heart disease (CHD). In this study, however, some subjects died from causes other than CHD or were censored due to loss to follow-up. Because the treatments were targeted to reduce CHD events, the investigators were interested in the effect of the treatment on CHD events in the absence of causes of death or events other than CHD. In this paper, we present a method for estimating treatment group-specific marginal survival curves of time-to-event data in the presence of dependent competing risks. The proposed method is a straightforward extension of the Inverse Probability of Censoring Weighted (IPCW) method to settings with more than one reason for censoring. The results of our analysis showed that the IPCW marginal incidence for CHD was almost the same as the lower bound for which subjects with competing events were assumed to be censored at the end of all follow-up. This result provided reassurance that the results in KLIS were robust to competing risks. Copyright © 2007 John Wiley & Sons, Ltd.

Keywords: *survival analysis; Kaplan–Meier estimator; Cox proportional hazards model; inverse weighting methods; cause of failure; competing risks*

1. INTRODUCTION

A common problem encountered in medical studies is competing risks, which are events that remove a subject from being at risk for the

*Correspondence to: Yutaka Matsuyama, Department of Biostatistics, School of Health Sciences and Nursing, University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
†E-mail: matuyama@epistat.m.u-tokyo.ac.jp

Pharmaceutical STATISTICS

outcome under investigation. The competing event may be the withdrawal of the subject from the study (for whatever reason), death from some cause other than the one of interest, or any eventuality that precludes the main event of interest from occurring. The Kyushu Lipid Intervention Study (KLIS) [1, 2], which is described in detail in Section 2, was a clinical study for hypercholesterolemia, where treatment with pravastatin (an HMG-CoA reductase inhibitor) was compared with conventional treatment in Japanese men aged 45–74 years. The primary endpoint was time to events of coronary heart disease (CHD). In this study, some subjects died from causes other than CHD such as cerebral infarction or cancer, and some subjects were lost to follow-up. Because the treatments were targeted to reduce CHD events, the investigators were not interested in the effect that treatment may have on competing events; rather, they were mainly interested in the effect of treatment on CHD events in the absence of causes of death or events other than CHD. This classical competing risks problem requires inference of the marginal distribution of time to CHD events [3, 4].

For the classical competing risks problems, Tsiatis [5] showed that the marginal survival function is not identifiable from observable data without additional assumptions. Independence of the latent failure times is one common assumption that would resolve the problem of identifiability and permit estimation of the marginal survival function using the product-limit estimator of Kaplan and Meier [6]. The assumption of independence, however, can never be verified from observed data and often may not be justified in practical settings. For example, in the KLIS data, one would expect that the subjects who have had a cerebral infarction would have a higher probability of CHD events due to the similarities of prognostic factors between the two events [7]. This kind of competing risk is obviously dependent on the event of interest. The Kaplan–Meier estimator under the assumption of independence will be inconsistent in the presence of dependent competing risks.

To address the issue of nonidentifiability, Peterson [8] provided upper and lower bounds on the marginal survival time as a function of the observed data. Peterson upper and lower bounds correspond to the extreme scenarios where censored subjects are assumed to never experience the event of interest or to experience the event immediately after censoring, respectively. In dealing with dependent competing risks, several authors have proposed bounds for the marginal survival function which are much tighter than those of Peterson [9–11]. These approaches impose certain dependency structures on the latent failure times which do not restrict the observed data, but allow for the identification of the marginal survival function under the knowledge of assumed dependency structures.

Recently, Robins and colleagues proposed the Inverse Probability of Censoring Weighted (IPCW) method for the analysis of data with informative censoring [12–15]. The IPCW estimator is a weighted version of the Kaplan–Meier and Cox partial likelihood estimators, in which weights are proportional to the inverse of the conditional probability of surviving uncensored [12, 14]. The IPCW method can be used to correct for bias due to dependent censoring when the dependent censoring can be explained by measured prognostic factors.

In this paper, we extend the IPCW approaches [14] to settings with more than one reason for censoring. In order to construct the IPCW marginal survival estimator, the cause-specific hazards of censoring are modeled by the proportional hazards model, in which the treatment group-specific baseline hazard and parameters are assumed for each reason of censoring. The next section describes the motivating study, the KLIS data. In Section 3, we introduce notation and describe the IPCW methods under competing risks. In Section 4, we apply the proposed methods to the KLIS data. In Section 5, we conclude with a discussion.

2. KLIS DATA

We briefly describe the motivating study and the data (KLIS). Full details on the design, conduct,

DOI: 10.1002/pst

Table I. Types and numbers of events in the KLIS data.

Types of events	Pravastatin treatment		Conventional treatment	
	Number	(%)	Number	(%)
CHD	65	2.9	48	2.9
No events at the end of follow-up	2033	91.6	1479	90.5
Loss to follow-up	53	2.4	44	2.7
Death due to causes other than CHD	68	3.1	63	3.9
Total	2219	100	1634	100

and main clinical results have been reported previously [1, 2]. A total of 5640 male patients aged 45–74 years with a serum total cholesterol level of $\geqslant 220$ mg/dl (5.69 mmol/l) in the pretest period were recruited by 902 physicians in Kyushu District from May 1990 to September 1993. Excluded from enrollment were those who withdrew their informed consent voluntarily; those who did not contact with study physicians thereafter; those with a history of myocardial infarction, coronary bypass surgery, coronary angioplasty, cerebral hemorrhage, or cerebral infarction; those with a high-density lipoprotein (HDL) cholesterol level of $\geqslant 80$ mg/dl; and those who had other specified life-limiting conditions such as renal failure or liver disease. The study subjects consisted of 3853 male patients.

Each study physician was instructed to allocate patients to either pravastatin treatment or conventional treatment randomly as specified in a sealed envelope, but some participating physicians did not necessarily follow the instruction of random assignment [1]. Therefore, the KLIS was essentially an observational study. In the protocol, pravastatin was prescribed at a dosage of 10–20 mg per day, which was an officially approved dose in Japan, and the conventional treatment included dietary and/or exercise therapy and medication with hypolipidemic drugs other than probucol, bezafibrate, and statins.

The primary endpoint was CHD events comprised of fatal and nonfatal myocardial infarction, coronary artery surgery, coronary angioplasty, cardiac death, and sudden and unexpected death. The endpoint and adverse effects committee regularly reviewed all possible cases of any event on the basis of the abstracts of medical records reported by the study physicians, laboratory data, and, if available, serial electrocardiograms. An underlying cause of death was classified and coded in accordance with the 9th revision of the ICD (International Classification Disease), based on a death certificate and/or a report abstract of the medical record. Follow-ups were continued until the end of 1997. From January to May 1998, an endpoint survey was carried out to make a full ascertainment of CHD events, other events such as cerebral infarction, and deaths from all causes until the end of 1997. Each study physician reconfirmed the summary information regarding outcome data, which was prepared by the study office based on the returned follow-up report forms. Study physicians made direct contact with patients who had ceased to visit the physicians by telephone and mail. Table I shows the types and numbers of events in each treatment group. The events were divided into four categories: CHD events; no events at the end of follow-up; loss to follow-up; and death due to causes other than CHD events, such as cerebral infarction, cancer, or suicide. The last two categories were regarded as competing events. The proportions of patients with competing events in the conventional treatment group were slightly larger than those in the pravastatin treatment group.

Figure 1 shows the Peterson bounds for the marginal incidence proportion of CHD events in each treatment group. Peterson [8] originally provided one extreme scenario as 'never experience the event of interest', that is, the unobserved event

DOI: 10.1002/pst

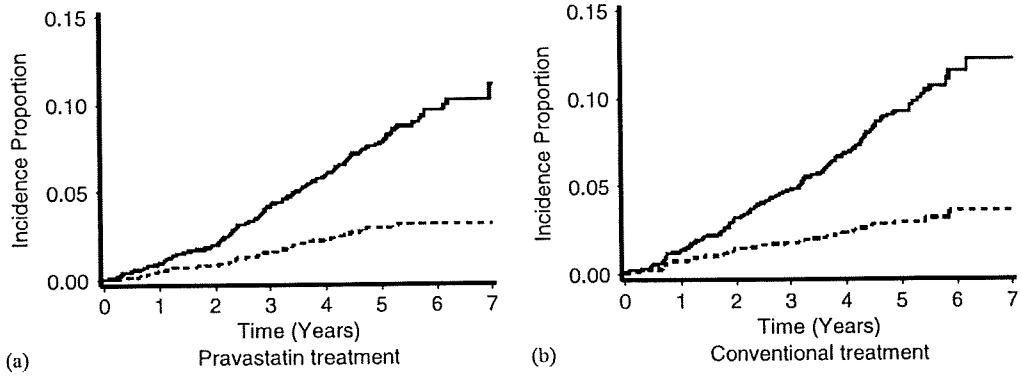


Figure 1. Peterson bounds for the marginal incidence proportion of CHD events in each treatment group. The solid line is one extreme scenario, in which subjects with competing events were assumed to experience CHD immediately after censoring. The dashed line is the other extreme scenario, in which subjects with competing events were assumed to be censored at the end of all follow-up.

Table II. The distributions of event times (years) of CHD and competing events.

Treatment group		Types of events			
		CHD	Loss to follow-up	Death due to causes other than CHD	All competing events*
Pravastatin	Number	65	53	68	121
	Mean	2.76	4.52	3.14	3.74
	Median	2.83	4.65	2.91	3.89
	Quartiles	1.67,3.74	3.99,5.57	2.09,4.15	2.34,5.03
	Range	0.19,5.30	0.16,7.46	0.22,7.25	0.16,7.46
Conventional	Number	48	44	63	107
	Mean	2.42	4.33	3.27	3.70
	Median	2.12	4.66	3.23	4.05
	Quartiles	1.00,3.73	4.08,4.97	2.15,4.33	2.55,4.71
	Range	0.05,5.86	0.31,6.65	0.71,7.29	0.31,7.29

*All competing events are 'loss to follow-up' plus 'death due to causes other than CHD'.

times are infinite. We regarded such competing events as censored at the end of all follow-up (the end of 1997). The usual treatment group-specific Kaplan–Meier estimates that ignore the competing risks, that is, the incidence proportion curves that censor all competing events at their event times, were almost the same as the lower bounds in each group. These observations were due to the fact that in each group most CHD events were likely to occur before the competing events, as shown in Table II. In this paper, we want to estimate the marginal incidence proportion in each treatment group, which will lie in between these two extreme scenarios.

3. IPCW METHODS

3.1. Assumption of no unmeasured confounders for censoring

Let T_i and C_i be the potential failure (occurrence of CHD events) time and the potential censoring time for subject i ($i = 1, 2, \ldots, n$), respectively. C_i

DOI: 10.1002/pst

is the minimum of C_{ij} $(j = 1, 2, 3)$, where C_{i1} denotes a death time due to causes other than CHD, C_{i2} denotes a censoring time of loss to follow-up, and C_{i3} denotes a censoring time at the end of follow-up. Censoring at the end of follow-up was not considered dependent censoring, because there was a fixed known calendar date at which the follow-up of all subjects ended (31/12/1997 for the KLIS data); therefore, C_{i3} was known for all subjects. The observable data are n i.i.d. copies of $X = \min(T, C_1, C_2, C_3)$, type of event J $(j = 0$, if CHD events are observed), treatment group indicator variable R ($R = 1$ if pravastatin treatment group, and $R = 0$ if conventional treatment group), and covariate history $\overline{\mathbf{V}}_X$, where $\overline{\mathbf{V}}_t = \{\mathbf{V}_s : 0 \leqslant s \leqslant t\}$, and $\mathbf{V}_s$ is a vector of possibly time-dependent prognostic factors for T recorded at time s.

In order to identify the marginal survival, we assume the following relation in the censoring process:

$$\lambda_{C_j}(t|R, \overline{\mathbf{V}}_t, T, T > t) = \lambda_{C_j}(t|R, \overline{\mathbf{V}}_t, T > t) \qquad (1)$$

where $j = 1, 2$ and $\lambda_{C_j}(t \mid (\cdot), T > t)$ is the cause-specific hazard of censoring at time t given both $X = \min(T, C_1, C_2, C_3)$ exceeds at t and the information in $(\cdot)$. This assumption means that, conditional on the treatment group and on the recorded history until time t, the cause-specific hazard of censoring C_j at time t does not depend on the possibly unobserved CHD event time T. We also assume an additional conditional independence assumption that the competing events C_j are independent of each other given the treatment group and the covariate history. This fundamental assumption is called 'no unmeasured confounders for censoring' [16] and is equivalent to a sequential version of Rosenbaum and Rubin's strong ignorability assumption [17]. The assumption specifies that, among subjects with the same recorded past, the population of subjects censored due to each specific cause at time t has the same distribution of the outcome of interest as that of the population of uncensored subjects at time t. The assumption will be satisfied, in particular, when the censoring process is ignorable or MAR (missing at random) in the terminology of missing data analysis. In practice, we would not expect this assumption to be precisely true, but given a rich collection of prognostic factors recorded in $\overline{\mathbf{V}}_t$, it may well be approximately true.

3.2. IPCW marginal survival time

The IPCW approach is to regard subjects with competing events as dependently censored the first time a subject either died or was censored by loss to follow-up. To correct for dependent censoring, we need to estimate the treatment group-specific hazards of censoring conditional on time-dependent prognostic factors for CHD [14]. The time-dependent Cox proportional hazards model for censoring is used for the right-hand side of equation (1),

$$\lambda_{C_j}(t|R, \overline{\mathbf{V}}_t, T > t) = \lambda_{0RC_j}(t) \exp(\alpha_{RC_j} \overline{\mathbf{V}}_t) \qquad (2)$$

where for each reason of censoring j $(j = 1, 2)$, the treatment group-specific baseline hazard $\lambda_{0RC_j}(t)$ and the treatment group-specific regression parameters α_{RC_j} are assumed, because both the baseline hazard and covariate effects may depend on treatment group. For estimating the hazard of a particular censoring type conditional on covariates, CHD events and other censoring types are censored at their event times.

Under the assumption of no unmeasured confounders for censoring (1) and the proportional hazards model for cause-specific hazards of censoring (2), the conditional probability of being uncensored for subject i is provided by the following time-dependent extension of the Kaplan–Meier estimator for censoring j,

$$\hat{K}_{ij}(t) = \prod_{u; X_u < t, \sigma_{uj} = 1, R_u = R_i} \exp[-\hat{\lambda}_{0RC_j}(X_u) \exp(\hat{\alpha}_{RC_j} \overline{\mathbf{V}}_{iX_u})]$$

where $\hat{\lambda}_{0RC_j}(X_u) = \sigma_{uj} / \sum_{i=1}^{n} \exp(\hat{\alpha}_{RC_j} \overline{\mathbf{V}}_{iX_u}) Y_i(X_u) I(R_i = R)$ is the Breslow estimator of the baseline hazard function for censoring j in treatment group R, and $Y_i(t)$ takes the value of one if subject i is at risk at time t, and zero otherwise. σ_{uj} takes the value of one if the subject is censored for reason j, and zero otherwise. For any proposition A, $I(A)$ equals one if A is true and zero otherwise.

DOI: 10.1002/pst

Pharmaceutical STATISTICS

The IPCW estimator is different from the ordinary estimator by weighting the contribution of a subject at risk by the inverse of the conditional probability of having remained uncensored. Using the above estimator of uncensored probability, the contribution of a subject at risk at time t is weighted by the inverse of an estimate of the conditional probability of having remained uncensored for both reasons until time t, $\hat{W}_i \times (t) = (1/\hat{K}_{i1}(t)) \times (1/\hat{K}_{i2}(t))$. Here, we assume that the conditional probabilities are bounded away from zero with probability 1 for each subject i, that is, $\hat{K}_{ij}(t) > 0$. This assumption will be satisfied unless their conditional probabilities are structural zero, that is, $\hat{K}_{ij}(t) = 0$ for some values of $\overline{\mathbf{V}}_t$. Under this assumption, the IPCW Kaplan–Meier estimator of the treatment group-specific marginal survival of not having CHD events through time t is

$$\hat{S}_T(t|R) = \prod_{\{i;X_i<t\}} \left\{ 1 - \frac{\delta_i \hat{W}_i(X_i) I(R_i = R)}{\sum_{u=1}^{n} Y_u(X_i) \hat{W}_u(X_i) I(R_u = R)} \right\} \tag{3}$$

where δ_i is the failure time indicator that takes the value of one if the subject failed and zero if the subject is censored. This IPCW Kaplan–Meier estimator for CHD events in treatment group R differs from the ordinary Kaplan–Meier estimator in that the contribution of a subject at any time X_i is weighted by the subject-specific weight $\hat{W}_i(X_i)$. In the IPCW estimator (3), the quantity, $\delta_i \hat{W}_i(X_i) I(R_i = R)$, estimates the number of subjects in treatment group R who would have been observed to fail at time X_i in the absence of any competing event, while the quantity, $\sum_{u=1}^{n} Y_u(X_i) \hat{W}_u(X_i) I(R_u = R)$, estimates the number of subjects who would have been alive and at risk at time X_i in the absence of any competing risk. Thus, the ratio estimates the hazard of CHD event at X_i in the absence of competing event; it follows that (3) estimates the probability $S_T(t|R)$ of surviving without failure (i.e. of remaining CHD-free) until time t in the absence of competing event. When (1) and (2) are true, Robins [12] proves that under mild regularity conditions, the IPCW estimator (3) gives a consistent estimator of our target causal estimand $S_T(t|R)$, that is, the marginal survival function. Inverse probability weighted estimators have been previously considered by Horvitz and Thompson [18] in the sample survey literature. Satten and Datta [19] give an elementary discussion of the IPCW estimators.

If the IPCW estimate (3) is close to the Peterson lower bound, we will see that the competing events are related to the unobserved CHD events and that these dependent competing risks can be explained by the covariates $\overline{\mathbf{V}}_t$ included in the analysis based on (2). On the other hand, if the IPCW estimate (3) is close to the Peterson upper bound, we will see that the competing events are not dependent ones under the assumption (1) that all important covariates were included in the covariate history based on (2). Therefore, it is important to compare the IPCW estimate (3) with the bounds in order to evaluate the degree and the direction of the selection bias.

3.3. Comparison of the IPCW marginal survival time

We used the Cox proportional hazards model to compare the marginal distribution between the two treatment groups. The model is

$$\lambda_T(t|R) = \lambda_0(t) \exp(\beta R) \tag{4}$$

where $\lambda_T(t|R)$ is the hazard of CHD events at time t in treatment group R. The IPCW Cox partial likelihood score $U(\beta)$ for β differs from the ordinary Cox partial likelihood score in that the contribution of the subject u at risk at time X_i is weighted by $\hat{W}_u(X_i)$, that is,

$$U(\beta) = \sum_{i=1}^{n} \delta_i \hat{W}_i(X_i) \times \left\{ R_i - \frac{\sum_{u=1}^{n} Y_u(X_i) \hat{W}_u(X_i) R_u \exp(\beta R_u)}{\sum_{u=1}^{n} Y_u(X_i) \hat{W}_u(X_i) \exp(\beta R_u)} \right\} \tag{5}$$

If (1) and (2) are correct, Robins [12] proves that under mild regularity conditions, the weighted estimating equations $U(\beta) = 0$ gives a consistent and asymptotically normal estimator of the parameter β.

DOI: 10.1002/pst

The use of individual weights induces within-subject correlation and we must take this correlation into consideration in the calculation of variance. In the calculation of a confidence interval, we used the robust variance estimate [20, 21]. It provides a conservative confidence interval for the parameter of interest, that is, the 95% Wald confidence interval calculated as $\beta \pm 1.96\times$ (robust standard error), which is guaranteed to cover the true value of β at least 95% of the time in large samples [21, 22]. We programmed the above procedure to obtain the IPCW estimate using SAS/IML procedure.

3.4. Adjustment of confounding by the IPTW method

In comparative studies, where investigators do not control treatment assignment, the directly estimated treatment effect can be strongly affected by confounding. This implies that in the KLIS data we cannot directly use the weighted log-rank-test (5) to compare the IPCW marginal incidence between treatment groups. There has been an enormous amount of work devoted to analytic adjustments for confounding. A new class of causal models called marginal structural models (MSMs) has recently been proposed [22, 23] to estimate the causal effect of treatment from observational data. In MSMs, the parameters are consistently estimated by the Inverse Probability of Treatment Weighted (IPTW) method. Here, we briefly describe the rationale for the method in the special case of a binary point treatment such as the KLIS data. A formal mathematical definition of MSMs using the counterfactual outcomes has been provided by Robins [23].

We consider the association model (4). In this subsection, we assume that there is no censoring, so the failure time T is observed on each subject. If a treatment assignment is completely at random and noncompliance is absent, the probability of receiving a treatment will be independent of both measured and unmeasured baseline prognostic factors, that is, there is no confounding. In this case, assuming that the association model (4) is correct, β has a causal interpretation, because association implies causation in the absence of confounding. This situation is called that a treatment is 'causally exogenous' [22, 23]. On the other hand, if the probability of receiving a treatment is independent of only measured baseline prognostic factors, a treatment is said to be 'statistically exogenous' [22,23]. It must be noted that the fact that a treatment is statistically exogenous does not imply that it is causally exogenous, because unmeasured confounders may predict the probability of receiving a treatment. We can empirically test whether a treatment is statistically exogenous, but not whether it is causally exogenous.

Suppose that we can correctly model the probability of receiving a treatment as a function of measured baseline prognostic factors $\mathbf{V}_0$. We could then quantify the degree to which the treatment is not statistically exogenous by the quantity,

$$W_{IPTW} = \frac{\Pr(R)}{\Pr(R \mid \mathbf{V}_0)}$$

where the denominator is the probability that a subject received his or her own observed treatment given measured prognostic factors $\mathbf{V}_0$, while the numerator is the unconditional probability that a subject received his or her own observed treatment. The numerator and denominator of W_{IPTW} are equal, if the treatment is statistically exogenous.

When the treatment is not statistically exogenous, we consider estimating β in the association model (4) by a weighted Cox regression in which a subject is given the weight W_{IPTW}. Weighting creates a pseudo-population where each subject is copied W_{IPTW} times. In this pseudo-population, the treatment is statistically exogenous and thus causally exogenous under the assumption of no unmeasured confounders. The weighted Cox regression estimator is called an IPTW estimator [22, 23]. As given in the Appendix 1 of Robins *et al.* [22], in a simple stratified point treatment analysis, the IPTW estimator is identical to the standardized estimator with the total group as the standard population. Hence, the MSM is interpretable as a nonparametric multivariate standar-

DOI: 10.1002/pst

dization method [24]. The weighted Cox regression to obtain the IPTW estimate can be performed with SAS/PHREG procedure using the 'WEIGHT' statement.

4. ANALYSIS OF KLIS DATA

To estimate the subject-specific weight $\hat{W}_i(X_i)$, we used five time-dependent factors as well as 12 baseline factors in the time-dependent Cox proportional hazards model for censoring (2). Prognostic factors at baseline were age (categorized as $-<49$ (reference category), 50–54, 55–59, 60–64, 65–69, and 70–74, with dummy variables for this categorization), serum HDL cholesterol (categorized as $-<40$ (reference category), $40-<48$, $48-<57$, and 57−mg/dl), serum LDL cholesterol (mg/dl), total cholesterol (categorized as $-<235$ (reference category), $235-<246$, $246-<262$, and 262−mg/dl), triglycerides (mg/dl), body mass index (BMI, categorized as $-<22.0$ (reference category), $22.0-<24.0$, $24.0-<25.9$, and $25.9-\text{kg/m}^2$), current smoking (dichotomous), diabetes mellitus (dichotomous), daily alcohol use (dichotomous), prior use of lipid-lowering drugs (dichotomous), hypertension (dichotomous), and angina pectoris (dichotomous). Time-dependent prognostic factors were serum HDL cholesterol, serum LDL cholesterol, total cholesterol, triglycerides, and the occurrence of cerebral infarction measured at six, 12, 24, 36, 48 and 60 months after a subject entered this trial. For these five time-dependent factors, the most recent recorded values were included as covariates in the prediction model (2) for the conditional probabilities of having remained uncensored. All these variables are clinically important prognostic factors for CHD events. The variable-selection procedures to reduce these variables to a relevant subset were not used, because it is important to include as many prognostic factors both for CHD events and for censoring as possible for the validity of our analysis, that is, the assumption (1) of 'no unmeasured confounders for censoring '.

Figure 2 shows the IPCW marginal incidence proportion of CHD events in each treatment group. In each group, the upper solid line is the Peterson bound, in which subjects with competing events were assumed to experience CHD immediately after censoring. In both treatment groups, the marginal incidence for CHD was almost the same as the lower bound in which subjects with competing events were assumed to be censored at the end of all follow-up. Therefore, there is little evidence of dependent competing risks in the KLIS data.

Because the KLIS was an observational study, we adjusted the baseline confounding by the IPTW method described in Section 3.4. We modeled the probability that a subject received the pravastatin

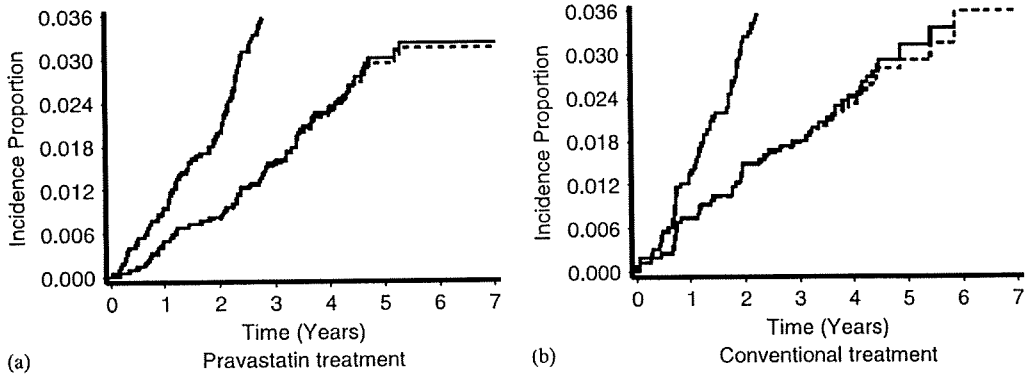


Figure 2. The IPCW marginal incidence proportion of CHD events in each treatment group. In each group, the upper solid line is the Peterson upper bound and the dashed line is the lower bound.

treatment using logistic regression with the 12 baseline factors described above as explanatory variables. From this logistic regression model, estimates of the subject-specific weight, $\hat{W}_i^*$, the inverse of the conditional probability of receiving his or her own observed treatment, were obtained. The subject-specific weight $\hat{W}_i(X_i) \times \hat{W}_i^*$ was used instead of $\hat{W}_i(X_i)$ or $\hat{W}_u(X_i)$ in the weighted score function (5). This weight is the inverse of the probability that a subject would have his or her own observed treatment and uncensored history for both reasons through time t. Figure 3 shows the results. A marginal treatment effect (hazard ratio = 0.78; 95% Wald confidence interval: 0.51–1.18) was observed after adjustments for baseline confounding as well as competing risks.

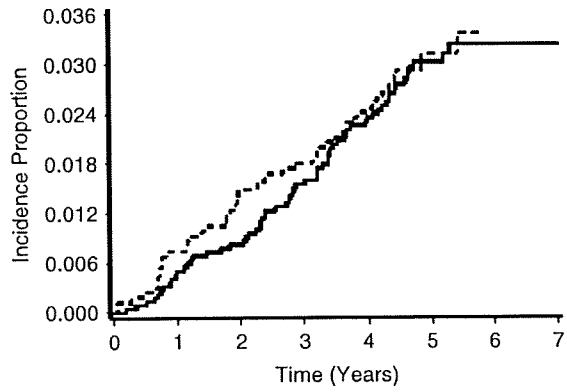


Figure 3. Comparison of the IPCW marginal incidence proportion of CHD events between treatment groups. Baseline confounding was adjusted by the IPTW method. The solid line is the pravastatin treatment group and the dashed line is the conventional treatment group.

Table III shows a comparison of hazard ratio and the 95% confidence interval under five adjustment methods. First analysis compared treatment effect without taking account of both baseline confounding and competing risks, in which competing events were assumed to be censored at their event times (hazard ratio = 0.97 ; 95% CI: 0.67–1.41). Second analysis compared treatment effect with an adjustment by the IPCW method taking account of only competing risks using the weighted log-rank-test (5) (hazard ratio = 0.94; 95% CI: 0.65–1.37). Third analysis compared treatment effect with an adjustment by the IPTW method taking account of only baseline confounding using the estimates of the subject-specific weight, $\hat{W}_i^*$, in which competing events were assumed to be censored at their event times (hazard ratio = 0.80; 95% CI: 0.53–1.20). Fourth analysis compared treatment effect with an adjustment by both the IPCW and the IPTW method, that is, Figure 3 (hazard ratio = 0.78; 95% CI: 0.51–1.18). A slightly stronger evidence of baseline confounding was observed with the crude result biased toward the null. However, the confidence intervals of the hazard ratio using different adjustment methods overlapped with each other. This might be due to the fact that the distributions of competing events were similar between treatment groups (Table II) and the KLIS was originally started as a randomized clinical trial [1, 2].

Final analysis compared treatment effect with an adjustment by the Cox regression models including all baseline covariates as the linear predictors and assuming all competing events to be censored at their event times (hazard ratio = 0.85; 95% CI: 0.57–1.27). It must be noted that,

Table III. Comparison of treatment effect.

No.	Analysis method	Hazard ratio	95% confidence interval
1*	No adjustments	0.97	0.67–1.41
2	Adjustment by the IPCW method	0.94	0.65–1.37
3*	Adjustment by the IPTW method	0.80	0.53–1.20
4	Adjustments by both the IPCW and IPTW methods	0.78	0.51–1.18
5*	Adjustment by the Cox regression model	0.85	0.57–1.27

*Competing events were assumed to be censored at their event times. In analysis No. 5, all baseline confounders were included as the linear predictors in the Cox regression model.

DOI: 10.1002/pst