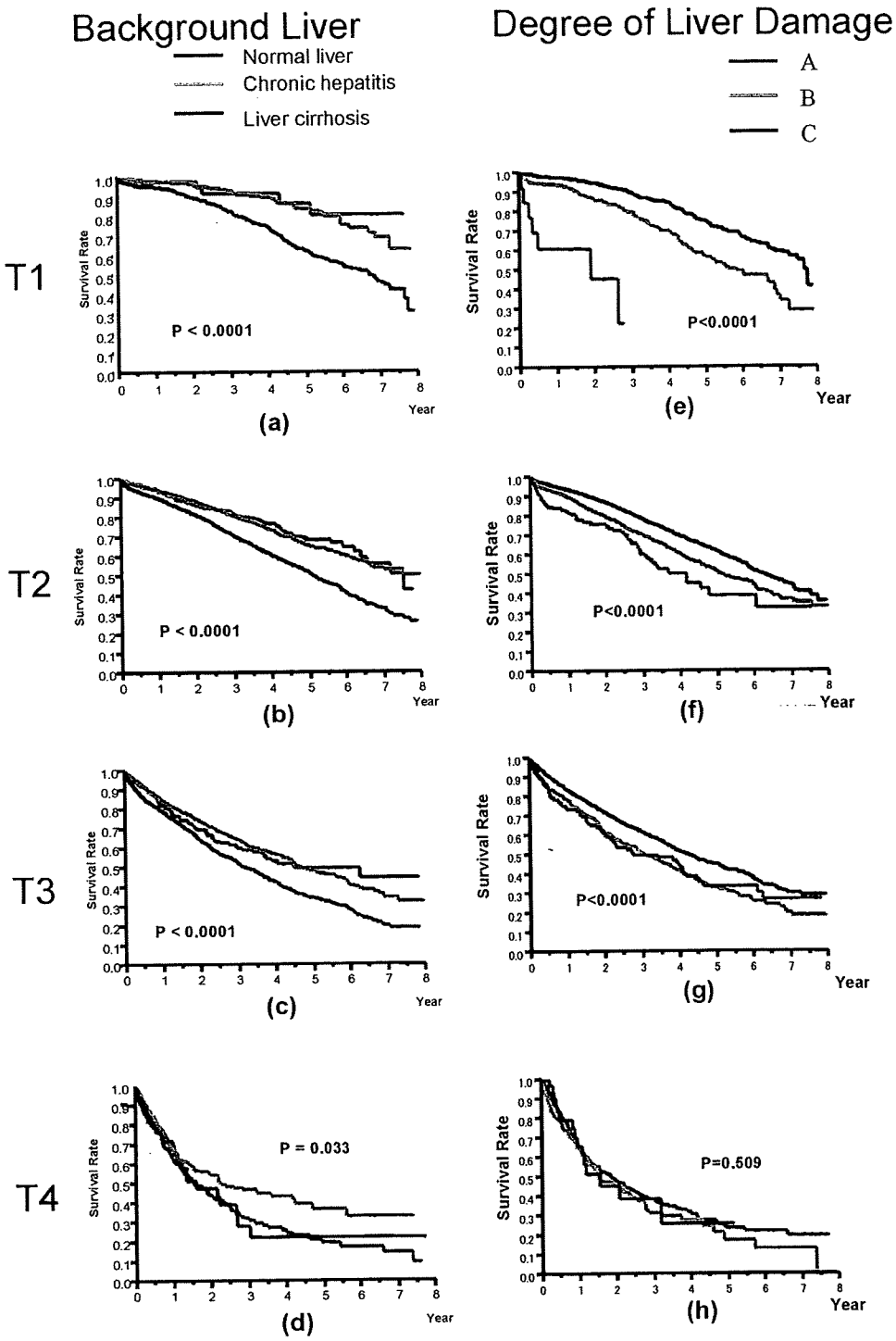

FIGURE 6. Influence of liver cirrhosis, chronic hepatitis (a, b, c, d) and the degree of liver damage (e, f, g, h) in T1, T2, T3, and T4.

third edition of LCSGJ.[26] This system required information on the growth pattern (single or multiple), size (≤2 cm or >2 cm), vascular invasion, and location (unilateral or bilateral) to determine the stage. It has been suggested that this system does not stratify patients adequately with respect to prognosis, and might be unnecessarily complex.[12,27–30] Izumi et al modified this system by eliminating the size and location (unilateral or bilateral).[27] The Izumi-TNM system was developed based on an analysis of 104 patients who underwent hepatic resection at a single center. Its prognostic value was demonstrated in 53 patients who received hepatectomy at an Italian center.[29,30] Influenced by this system, Vauthey et al[13]

TABLE 8. Predictive Model Based on the T Classification

Parameter	Estimate	Standard Error	P	Hazard Ratio	95% CI
LCSGJ-T					
T2 vs.T1	0.562	0.081	0.001	1.75	1.50–2.06
T3 vs. T1	1.091	0.085	0.001	2.98	2.52–3.52
T4 vs. T1	1.749	0.112	0.001	5.75	4.62–7.16
AJCC-T					
T2 vs. T1	0.497	0.054	0.001	1.64	1.48–1.83
T3 vs. T1	1.126	0.067	0.001	3.08	2.71–3.51

proposed a scoring system that included size (≤5 cm or >5 cm), vascular invasion, and growth pattern (single or multiple): patients with 0, 1, 2, and all 3 of these factors were assigned to T1, T2, T3, and T4, respectively.[31] They applied this T-stage to 97 patients and demonstrated its ability to stratify patients. It was also validated in 323 patients who received curative hepatic resection in Taiwan.[32] In 2002, Vauthey et al developed a simplified staging system based on a survival analysis of 557 patients who received hepatic resection at 4 hepatobiliary centers in the United States, France, and Japan.[12] It was adopted as the AJCC Cancer Staging Manual, 6th edition, in 2002.[13]

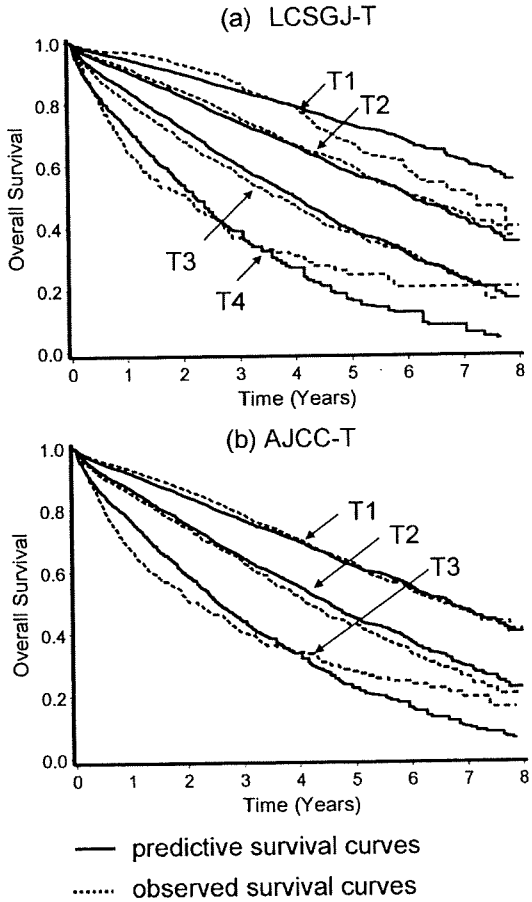

FIGURE 7. Comparison of the predictive survival curves (solid line) and the observed survival curves (broken line) in the validation sample. Prediction based on the LCSGJ-T classification (a), and AJCC T classification (b).

The major differences between LCSGJ 4th TNM and AJCC/UICC 6th TNM are the cutoff value for tumor size and its application in prognostic classification (Fig. 1). In the revision from the fifth to sixth edition of the AJCC/UICC TNM staging, the cutoff value for tumor size in the prognostic classification was shifted from 2 cm to 5 cm, based on the results of Vauthey et al.[12] In their series, tumor size had no effect on survival in patients with no vascular invasion or microvascular invasion, and had a significant effect on survival only in patients with multiple HCCs.[12] The lack of a significant difference may be related to the small sample size. In the AJCC/UICC staging system, a single tumor without vascular invasion is assigned to T1, while a tumor with vascular invasion is assigned to T2 regardless of its size. In our series, the prognosis of patients with a single tumor without vascular invasion showed a wide range according to its size (Fig. 4b): patients with tumors of 2 cm or smaller showed a 5-year survival rate of 70%, while those with tumors of 10 cm or larger had a 5-year survival rate of 49%. The prognosis of patients with vascular invasion is also greatly influenced by the tumor size (Fig. 4d), and that of patients with multiple tumors gradually worsened with increasing tumor size in our data (Fig. 4c, e). Patients with tumors of 2 cm or smaller showed a significantly favorable prognosis regardless of vascular invasion or growth pattern (single or multiple) (Fig. 4b–e). Among tumors 2 cm or smaller, about 22% are early HCCs, which are defined as tumors that have Glisson's triad and show hypercellularity of over 2-fold with minimal cellular or nuclear atypia (Edmondson's grade 1).[33,34] These tumors showed a remarkably favorable prognosis: a 5-year survival rate of 93% was reported.[34] Ideally, patients with these tumors should be graded to the earliest stage, but the definition of these tumors requires a pathologic examination of a resected specimen. To simplify the diagnosis, a single HCC of 2 cm or smaller without vascular invasion should be graded as the earliest stage.

In the current study, the presence of liver cirrhosis in the background liver and the degree of liver damage[10,22] were also independent prognostic factors ($P < 0.0001$) (Table 5; Fig. 6). This factor significantly influenced the prognosis in the T1, T2, and T3 subsets. The influence of liver damage on survival decreases with T stage, until it is no longer significant for patients with T4 HCC (Fig. 6h). This may be a

TABLE 9. Comparison of Predictive Accuracy of Patient Survival Between LCSGJ and AJCC-T Classifications

Parameter	Estimate	Standard Error	*P*
AJCC-T	0.0085	0.0004	0.0001
LCSGJ-T	0.0068	0.0008	0.0001
Score of T classification	0.0061	0.0003	0.0001
LCSG vs. AJCC	−0.0017	0.00049	0.0007

consequence of the fact that tumor factors govern the prognosis of patients with advanced HCC and liver function plays an important role in patients with relatively early HCC. Thus, a scoring system that uniformly assigns the tumor stage and liver function, for example, the score proposed by CLIP or JIS, may be limited in its ability to stratify patients with an advanced score.[4,8,35,36] Accordingly, a staging system that is composed only of tumor factors and is additionally subclassified by liver cirrhosis in the background liver or liver function, like AJCC stage or LCSG stage, may be simple and suitable for patients who will undergo curative treatment such as hepatectomy, radiofrequency ablation, or liver transplantation.

A weakness of this LCSGJ staging system may be a result of the assumption of equal weight for growth pattern (single or multiple), size, and vascular or bile duct invasion. In our data, vascular or bile duct invasion had the strongest impact on survival: its relative risk (RR) was 1.36, followed by liver cirrhosis (RR 1.26), diameter (RR 1.21), alpha-fetoprotein (RR 1.20), and tumor number (RR 1.18) (Table 5). Therefore, patients with T2 or T3 tumor can be divided into 3 subgroups according to these factors, and it is logical to hypothesize that the subgroup with a more heavily weighted prognostic factor, ie, vascular or bile duct invasion, will have a worse prognosis. In our data, no survival differences were observed among patients with T2 tumor, although patients with T3 tumor were layered with regard to their prognosis: those with vascular or bile duct invasion showed a significantly worse prognosis. Interestingly, patients with T3 tumor accompanied by vascular or bile duct invasion had a longer survival than those with T4 tumor, and patients with T3 tumor that was not accompanied by vascular or bile duct invasion had a worse outcome than those with T2 tumor. Accordingly, the cohort of T3 tumor should be subdivided according to the presence of vascular or bile duct invasion.

The preferred treatment of small HCC (hepatic resection, radiofrequency ablation, or transplantation) has been controversial for many years.[1] In the AJCC T classification, 8457 of 12,534 patients (67.5%) were assigned to T1, while in the LCSGJ T classification 16.6% were assigned to T1 (Table 7). Of the 8457 patients with T1 HCC in the AJCC T classification, 2078 were classified as T1, 6308 were T2, and 71 were T3 in the LCSGJ T classification. As expected, patients with T1 tumor in AJCC T classification had a 5-year survival rate of 61%, which was similar to that of patients with T2 tumor in the LCSGJ T classification (58%). These results suggest that the LCSGJ-T classification may make it possible to classify patients with early-stage HCC more precisely than with AJCC T.

Both the LCSGJ-stage and the AJCC-stage were developed based on a survival analysis of patients who underwent hepatic resection. Thus, these staging systems are appropriate for patients who will undergo hepatic resection. The applicability of these surgical staging systems to other local therapies such as transplantation and radiofrequency ablation has been a matter of dispute. The previous version of the AJCC TNM classification was shown by Llovet et al to not have prognostic power in patients who were treated by liver transplantation,[37] although Peck-Radosavljevic et al showed that stage IVA predicted recurrence in these patients.[38] The current AJCC TNM system 6th edition has been shown to be a significant predictor of tumor recurrence, but not of survival in liver transplantation.[39] Machi et al studied 65 patients with unresectable HCC who underwent radiofrequency ablation and showed that TNM stage 6th edition was a significant predictor of survival.[40] The applicability of LCSGJ TNM stage to liver transplantation or radiofrequency ablation has not been fully evaluated. Thus, it may be preferable to apply these surgical staging systems only in patients who will receive liver resection.

All staging systems represent a compromise between simplicity and discriminatory ability. Although we tried to reduce the over-fitting bias by internal cross-validation, a further independent external validation of the LCSGJ-TNM stage is needed.

ACKNOWLEDGMENTS

The authors thank all members of the Liver Cancer Study Group of Japan who completed the questionnaires. The authors also thank Shigeki Arii, MD, PhD, from the Department of Hepato-Biliary-Pancreatic Surgery, Tokyo Medical and Dental University, Graduate School of Medicine, Tokyo, Japan; Takafumi Ichida, MD, PhD, from the Department of Gastroenterology, Juntendo University School of Medicine, Tokyo, Japan; Kiwamu Okita, MD, PhD, from the Department of Gastroenterology and Hepatology, Yamaguchi University School of Medicine, Ube, Japan; Masao Omata, MD, PhD, from the Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; Masamichi Kojiro, MD, PhD, from the Department of Pathology, Kurume University School of Medicine, Kurume, Japan; Yasuni Nakanuma, MD, PhD, from the Department of Human Pathology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan; and Kenichi Takayasu, MD, PhD, from the Department of Diagnostic Radiology, National Cancer Center Hospital, Tokyo, Japan, for their contributions to this study.

REFERENCES

1. Arii S, Yamaoka Y, Futagawa S, et al. Results of surgical and nonsurgical treatment for small-sized hepatocellular carcinomas: a retrospective and nationwide survey in Japan. The Liver Cancer Study Group of Japan. *Hepatology.* 2000;32:1224–1229.
2. Chen MS, Li JQ, Zheng Y, et al. A prospective randomized trial comparing percutaneous local ablative therapy and partial hepatectomy for small hepatocellular carcinoma. *Ann Surg.* 2006;243:321–328.
3. Okuda K, Ohtsuki T, Obata H, et al. Natural history of hepatocellular carcinoma and prognosis in relation to treatment: study of 850 patients. *Cancer.* 1985;56:918–928.

4. Anonymous. A new prognostic system for hepatocellular carcinoma: a retrospective study of 435 patients. The Cancer of the Liver Italian Program (CLIP) investigators. *Hepatology.* 1998;28:751–755.
5. Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. *Semin Liver Dis.* 1999;19:329–338.
6. Chevret S, Trinchet JC, Mathieu D, et al. A new prognostic classification for predicting survival in patients with hepatocellular carcinoma: Groupe d'Etude et de Traitement du Carcinome Hepatocellulaire. *J Hepatol.* 1999;31:133–141.
7. Leung TW, Tang AM, Zee B, et al. Construction of the Chinese University Prognostic Index for hepatocellular carcinoma and comparison with the TNM staging system, the Okuda staging system, and the Cancer of the Liver Italian Program staging system: a study based on 926 patients. *Cancer.* 2002;94:1760–1769.
8. Kudo M, Chung H, Osaki Y. Prognostic staging system for hepatocellular carcinoma (CLIP score): its value and limitations, and a proposal for a new staging system, the Japan Integrated Staging Score (JIS score). *J Gastroenterol.* 2003;38:207–215.
9. Henderson J, Sherman M, Tavill A, et al. AHPBA/AJCC consensus conference on staging of hepatocellular carcinoma: consensus statement. *HPB.* 2003; 5:243–250.
10. Liver Cancer Study Group of Japan. *General Rules for the Clinical and Pathological Study of Primary Liver Cancer,* 4th Japanese edition. Tokyo: Kanehara, 2000.
11. Liver Cancer Study Group of Japan. *General Rules for the Clinical and Pathological Study of Primary Liver Cancer,* 2nd English edition. Tokyo: Kanehara, 2003. [The 2000 4th Japanese edition corresponds to the 2003 2nd English edition.]
12. Vauthey JN, Lauwers GY, Esnaola NF, et al. Simplified staging for hepatocellular carcinoma. *J Clin Oncol.* 2002;20:1527–1536.
13. AJCC. *AJCC Cancer Staging Manual,* 6th ed. New York: Springer, 2002.
14. Makuuchi M, Belghiti J, Belli G, et al. IHPBA concordant classification of primary liver cancer: working group report. *J Hepatobiliary Pancreat Surg.* 2003;10:26–30.
15. Ueno S, Tanabe G, Nuruki K, et al. Prognostic performance of the new classification of primary liver cancer of Japan (4th edition) for patients with hepatocellular carcinoma: a validation analysis. *Hepatol Res.* 2002; 24:395–403.
16. Poon RT, Fan ST. Evaluation of the new AJCC/UICC staging system for hepatocellular carcinoma after hepatic resection in Chinese patients. *Surg Oncol Clin North Am.* 2003;12:35–50.
17. Anonymous. Primary liver cancers in Japan. *Cancer.* 1980;45:2663–2669.
18. Anonymous. Primary liver cancer in Japan: the Liver Cancer Study Group of Japan. *Cancer.* 1984;54:1747–1755.
19. Anonymous. Primary liver cancer in Japan, Sixth report: the Liver Cancer Study Group of Japan. *Cancer.* 1987;60:1400–1411.
20. Anonymous. Primary liver cancer in Japan: clinicopathologic features and results of surgical treatment. Liver Cancer Study Group of Japan. *Ann Surg.* 1990;211:277–287.
21. Anonymous. Predictive factors for long term prognosis after partial hepatectomy for patients with hepatocellular carcinoma in Japan: the Liver Cancer Study Group of Japan. *Cancer.* 1994;74:2772–2780.
22. Ikai I, Arii S, Kojiro M, et al. Reevaluation of prognostic factors for survival after liver resection in patients with hepatocellular carcinoma in a Japanese nationwide survey. *Cancer.* 2004;101:796–802.
23. Ikai I, Itai Y, Okita K, et al. Report of the 15th follow-up survey of primary liver cancer. *Hepatol Res.* 2004;28:21–29.
24. A Japanese study group for alcoholic liver disease: a new diagnostic criteria of alcoholic liver disease. Takada T, *Acta Hepatol Jpn.* 1993; 34:888–896.
25. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986;73:13–22.
26. Vauthey JN, Pawlik TM, Lauwers GY, et al. Critical evaluation of the different staging systems for hepatocellular carcinoma. *Br J Surg.* 2004;91:1072.
27. Izumi R, Shimizu K, Ii T, et al. Prognostic factors of hepatocellular carcinoma in patients undergoing hepatic resection. *Gastroenterology.* 1994;106:720–727.
28. Cance WG, Stewart AK, Menck HR. The National Cancer Data Base Report on treatment patterns for hepatocellular carcinomas: improved survival of surgically resected patients, 1985–1996. *Cancer.* 2000;88:912–920.
29. Staudacher C, Chiappa A, Biella F, et al. Validation of the modified TNM-Izumi classification for hepatocellular carcinoma. *Tumori.* 2000; 86:8–11.
30. Chiappa A, Zbar AP, Podda M, et al. Prognostic value of the modified TNM (Izumi) classification of hepatocellular carcinoma in 53 cirrhotic patients undergoing resection. *Hepatogastroenterology.* 2001; 48:229–234.
31. Vauthey JN, Klimstra D, Blumgart LH. A simplified staging system for hepatocellular carcinomas. *Gastroenterology.* 1995;108:617–618.
32. Lui WY, Chiu ST, Chiu JH, et al. Evaluation of a simplified staging system for prognosis of hepatocellular carcinoma. *J Formos Med Assoc.* 1999;98:248–253.
33. Kanai T, Hirohashi S, Upton MP, et al. Pathology of small hepatocellular carcinoma: a proposal for a new gross classification. *Cancer.* 1987;60:810–819.
34. Takayama T, Makuuchi M, Hirohashi S, et al. Early hepatocellular carcinoma as an entity with a high rate of surgical cure. *Hepatology.* 1998;28:1241–1246.
35. Anonymous. Prospective validation of the CLIP score: a new prognostic system for patients with cirrhosis and hepatocellular carcinoma. The Cancer of the Liver Italian Program (CLIP) Investigators. *Hepatology.* 2000;31:840–845.
36. Kudo M, Chung H, Haji S, et al. Validation of a new prognostic staging system for hepatocellular carcinoma: the JIS score compared with the CLIP score. *Hepatology.* 2004;40:1396–1405.
37. Llovet JM, Bruix J, Fuster J, et al. Liver transplantation for small hepatocellular carcinoma: the tumor-node-metastasis classification does not have prognostic power. *Hepatology.* 1998;27:1572–1577.
38. Peck-Radosavljevic M, Pidlich J, Bergmann M, et al. Preoperative TNM classification is a better prognostic indicator for recurrence of hepatocellular carcinoma after liver transplantation than albumin mRNA in peripheral blood: Liver Transplant Oncology Group. *J Hepatol.* 1998;28:497–503.
39. Zavaglia C, De Carlis L, Alberti AB, et al. Predictors of long-term survival after liver transplantation for hepatocellular carcinoma. *Am J Gastroenterol.* 2005;100:2708–2716.
40. Machi J, Bueno RS, Wong LL. Long-term follow-up outcome of patients undergoing radiofrequency ablation for unresectable hepatocellular carcinoma. *World J Surg.* 2005;29:1364–1373.

CLINICAL TRIALS ARTICLE

Clinical Trials 2007; 4: 318–328

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)

Mizuki Yoshida[a], Yutaka Matsuyama[b] and Yasuo Ohashi[b], for the MEGA Study Group

Background The MEGA study is a randomized controlled trial conducted in Japan to evaluate the primary preventive effect of pravastatin against coronary heart disease (CHD), in which 8214 subjects are randomized to diet or diet plus pravastatin. Pravastatin reduces the incidence of CHD (hazard ratio = 0.67; 95%CI: 0.49–0.91). In the MEGA study, in addition to the usual loss to follow-up cases, there is another problem of drop-outs due to the refusal of further follow-up at 5 years.

Purpose To estimate the treatment effect adjusting for some types of dependent censorings observed in the MEGA study and to assess the sensitivity of standard analysis results for these censoring cases.

Methods The proposed method is a straightforward extension of the inverse probability of censoring weighted (IPCW) method for settings with more than one reason for censoring, where the propensities for drop-outs are modeled separately for each reason. Simulation studies are also conducted to compare the properties of the IPCW estimate with the standard analysis assuming independent censorings.

Results Simulation studies show that the IPCW estimate can correct for selection bias due to dependent censoring that can be explained by measured factors, while the standard analysis is biased. Applying the proposed method to the MEGA study data, several prognostic factors are associated with the censoring processes, and after adjusting for these dependent censorings, slightly larger treatment effects for pravastatin are observed for both CHD (primary endpoint) and stroke (secondary endpoint) events.

Limitations The method developed is based on the fundamental assumption of sequentially ignorable censoring.

Conclusions Our proposed method provides a valuable approach for estimating treatment effect adjusting for several types of dependent censorings. Dependent censorings observed in the MEGA study did not cause a severe selection bias attributable to the covariates and the results from the standard analysis were robust in relation to the censorings. *Clinical Trials* 2007; 4: 318–328. http://ctj.sagepub.com

[a]Biometrics Division, The Clinical Service Provider, EPS Co., Ltd., 2-3-19 Koraku, Bunkyo-ku, Tokyo 112-0004, Japan
[b]Department of Biostatistics/Epidemiology and Preventive Health Sciences, School of Health Sciences and Nursing, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
Author for correspondence: Yutaka Matsuyama, Department of Biostatistics, School of Health Sciences and Nursing, University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-0033, Japan, Tel: +81-3-5841-3519, Fax: +81-3-5841-3527. E-mail: matuyama@epistat.m.u-tokyo.ac.jp
Grant Support: 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.

10.1177/1740774507081224

Introduction

The management of elevated cholesterol in the primary prevention group of adult Japanese (MEGA) study is a randomized controlled trial conducted in Japan to evaluate the primary preventive effect of a statin against coronary heart disease (CHD) in daily clinical practice [1]. In this prospective, randomized, open-labeled, blinded-endpoints (PROBE) design study, men and postmenopausal women aged 40–70 years with hypercholesterolemia (total cholesterol (TC) level: 220–270 (mg/dL)) and no history of CHD or stroke were randomized to diet (diet group) or diet plus pravastatin 10–20 mg daily (pravastatin group). The primary endpoint was the first occurrence of CHD, comprising fatal and nonfatal myocardial infarction, angina, cardiac and sudden death, and a coronary revascularization procedure. One of the secondary endpoints was the first occurrence of stroke.

Between February 1994 and March 1999, a total of 15 210 persons visiting an outpatient clinic were registered throughout Japan. Of the 15 210 subjects who met the inclusion criteria regardless of their TC level and who provided signed informed consent, 8214 who met the TC criterion were randomized to either diet or diet plus pravastatin treatment using the permuted block method with stratification according to gender, age, and medical institution. Of the randomized subjects, 382 were excluded; 94 withdrew their consent, 224 had exclusion criteria violation after randomization, and 64 had no recorded data after randomization. The remaining 7832 patients were analyzed (3966 diet group; 3866 pravastatin group). Follow-up was continued until March 2004. The incidence of CHD was significantly lower by 33% in the pravastatin group than in the diet group (hazard ratio = 0.67; 95%CI: 0.49–0.91; $p = 0.01$) [2].

A common problem encountered in any survival analysis is censoring data with a possible non-ignorable response mechanism. The response mechanism, which is the reason whether or not a response is obtained, is said to be non-ignorable if it depends on a subject's unobserved response [3]. If the censoring times are stochastically independent of survival time, the censoring is ignorable and standard survival analysis methods assuming independent censoring is valid. For example, an end-of-study censoring is completely determined by the enrollment time. If the survival does not change over time, such censoring time is independent of survival time. The assumption of independence, however, can never be verified from observed data and often may not be justified in practical settings. For example, one would suspect that drop-out subjects are different from the other subjects with respect to many background characteristics including the histories of prognostic factors such as lipid values. This type of drop-out may be dependent on the event of interest. The Kaplan–Meier estimator or the log-rank test under the assumption of independence will be inconsistent in the presence of dependent censorings [4].

In the MEGA study, although the follow-up period was initially scheduled for 5 years, based on the recommendation of the Data and Safety Monitoring Committee, the study was continued an additional 5 years to increase the number of events, and thus, patients who provided written consent at 5 years to continue the study were followed until the end of March 2004 [1,2]. Therefore, in addition to the usual loss to follow-up cases, there is another problem of drop-outs due to the refusal of further follow-up at 5 years. To ensure that the results in MEGA study are robust in relation to its censorings, it is important to assess the sensitivity of standard analysis results [2] to these possibly dependent drop-out cases.

Recently, Robins and colleagues proposed the inverse probability of censoring weighted (IPCW) method for the analysis of data with informative censoring [5–9]. The underlying idea of the IPCW method is to base estimation on the observed responses but weight them to account for the probability of remaining in the study. The propensities for drop-outs can be estimated as a function of the observed responses prior to drop-outs, and also as a function of the covariates and any additional variables or subject characteristics that are thought likely to predict drop-outs. 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 article, we extend the IPCW approaches for time-to-event data [7] to settings with more than one reason for censoring. To obtain the probability of remaining in the study, we use separate models for each drop-out process, because, in the MEGA study, one type of drop-out dominates at 5 years, and the other type dominates otherwise. This modeling strategy is important, because there is a possibility of important differences in the effect of various predictors on each separate type of drop-out, and causal interpretation of the IPCW estimates depends on the correct specification of the model for drop-out [7,9]. For the usual drop-out cases such as loss to follow-up, the cause-specific hazards of censoring are modeled by the time-dependent Cox proportional hazards model, in which the treatment group-specific baseline hazard and parameters are separately assumed in the two treatment groups. For the

drop-outs due to the refusal of follow-up, the probability of drop-outs at 5 years is modeled by the logistic regression model. These two estimated weights are combined in order to construct the IPCW Kaplan–Meier estimator and the IPCW log-rank statistic. The remainder of this article is divided into five sections: presenting the MEGA study data; describing the proposed IPCW methodology; presenting the simulation studies to evaluate the performance of the IPCW estimation method; presenting the analysis results of the MEGA study data; and finally, conclusion with some discussion.

MEGA study

We will briefly describe the MEGA study data. Full details on the design, conduct, and main clinical results have been reported [1,2]. Table 1 shows the baseline characteristics of the analyzed patients. There was no clinical difference between the two groups in baseline characteristics. Women accounted for 68.4% (5356 patients) of the study population. Mean body mass index (BMI) was 23.8 (kg/m^2). Mean TC, low-density lipoprotein cholesterol (LDL-C), and high-density lipoprotein cholesterol (HDL-C) levels were 242.6, 156.6, and 57.5 (mg/dL), respectively. Median triglyceride (TG) level was 127.5 (mg/dL). Of the study patients, 41.8 and 20.8% had hypertension and diabetes mellitus based on physician diagnosis, respectively.

After randomization, patients were followed at months 1, 3, and 6 and thereafter every 6 months. At each visit, data on treatment compliance, use of concomitant drugs, onset of events, occurrence of adverse events, and laboratory tests including serum lipids were collected by the investigators. Additionally, an ECG was obtained and evaluated annually. All endpoints were reviewed strictly by the blinded Endpoint Committee and additional information obtained from the physician as needed [1]. A total of 7832 patients were followed by 2658 physicians in 1320 hospitals. The follow-up period was 41195 person-years (mean follow-up period 5.3 years). Table 2 shows the types and numbers

Table 1 Baseline characteristics of analyzed 7832 patients

Characteristics	Diet group $N=3966$		Diet + pravastatin group $N=3866$	
Age (years), mean (SD)	58.4	(7.2)	58.2	(7.3)
Women, No. (%)	2718	(68.5)	2638	(68.2)
BMI (kg/m^2), mean (SD)	23.8	(3.0)	23.8	(3.1)
Current smoker, No. (%)	572	(14.4)	612	(15.8)
Current drinking, No. (%)	1183	(29.8)	1180	(30.5)
Hypercholesterolemia medication history, No. (%)	621	(15.7)	586	(15.2)
Hypertension, No. (%)	1664	(42.0)	1613	(41.7)
Diabetes, No. (%)	828	(20.9)	804	(20.8)
TC (mg/dL), mean (SD)	242.6	(12.1)	242.6	(12.0)
G (mg/dL), median (inter-quartile range)	127.5	(95.0–179.0)	127.4	(95.7–176.5)
HDL-C (mg/dL), mean (SD)	57.5	(15.1)	57.5	(14.8)
LDL-C (mg/dL), mean (SD)	156.5	(17.3)	156.7	(17.6)

SD: standard deviation; BMI: body mass index; TC: total cholesterol; TG: triglyceride; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol

Table 2 Types and numbers of events

Types of events	Diet group		Diet + pravastatin group	
	Number	(%)	Number	(%)
CHD	101	(2.5)	66	(1.7)
Loss to follow-up	546	(13.8)	594	(15.4)
Refusal of follow-up by patients	278	(7.0)	270	(7.0)
Refusal of follow-up by institutions	165	(4.2)	162	(4.2)
End-of-study censoring	2876	(72.5)	2774	(71.8)
Total	3966	(100)	3866	(100)

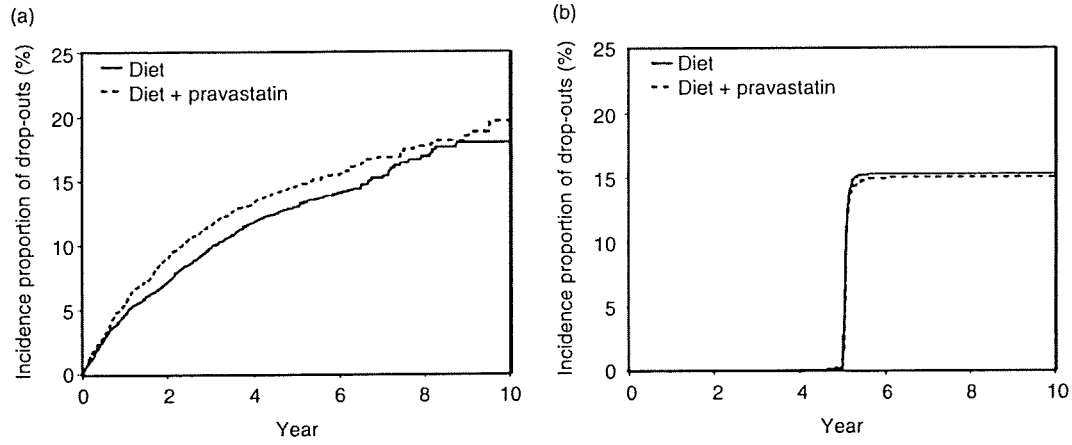

Figure 1 Incidence proportion for drop-outs (a) loss to follow-up; (b) refusal of follow-up by patients

of events in each treatment group. The events were divided into five categories: 1. CHD events; 2. loss to follow-up; 3. refusal of follow-up at 5 years by patients; 4. refusal of follow-up at 5 years by institutions; and 5. no events at the end of study. The withdrawal of informed consent occurring except at 5 years was included in the category of loss to follow-up (283 patients in diet group, 382 patients in pravastatin group). The refusal of follow-up at 5 years was divided into two categories, because, when obtaining the consent to continue the study, each Institutional Review Board (IRB) firstly made the decision regardless of the patient's intention. This institution-specific drop-out at 5 years was not related to the patient's medical histories and was thought to be an end-of-study censoring. These censorings at the end of study (refusal of follow-up by institutions and no events at the end of study) were not considered dependent censoring, because there was a fixed known calendar date at which the follow-up ended. Therefore, the second (Reason 1: loss to follow-up) and the third (Reason 2: refusal of follow-up by patients) categories were regarded as dependent censorings in this study.

Figure 1 shows the Kaplan–Meier curves for the event of drop-outs, Reasons 1 and 2, respectively. For Reason 1, more drop-outs were observed in the pravastatin group. For Reason 2, the times of drop-outs were distributed around 5 years after the start of follow-up and there was no distributional difference between the two treatment groups. Figure 2 shows the Kaplan–Meier curves for the CHD events that censored all drop-out cases at their event times. The incidence of CHD was significantly lower by 33% in the pravastatin group than in the diet group (hazard ratio = 0.67; 95%CI: 0.49–0.91; $p = 0.01$) [2].

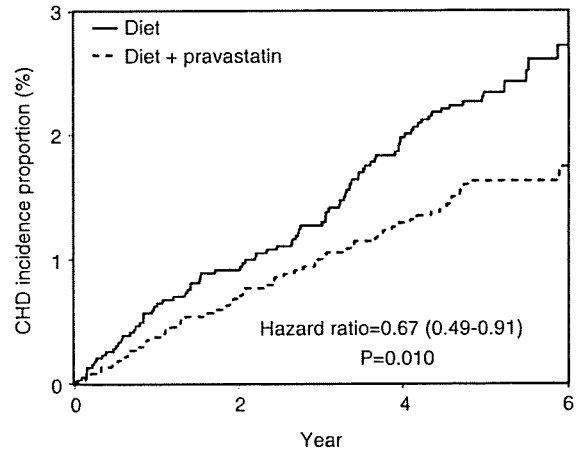

Figure 2 Incidence proportion for CHD events under the assumption of independent censoring

IPCW methods

Notation and 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 ($i = 1, \ldots, n$), respectively. C_i is the minimum of C_{ij} ($j = 1, 2, 3$), where C_{i1} denotes a censoring time of loss to follow-up (Reason 1), C_{i2} denotes a censoring time due to refusal of follow-up by patients (Reason 2), and C_{i3} denotes a censoring time at the end of study. 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 diet plus pravastatin group, and $R = 0$ if diet group), and

covariate history $\overline{V}_X$, where $\overline{V}_t = \{V_s: 0 \le s \le t\}$, and V_s is a vector of possibly time-dependent prognostic factors for T recorded at time s.

In order to identify the survival time in the presence of dependent censoring, we assume the following relation in the censoring process,

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

where $j = 1, 2$ and $\lambda_{C_j}(t|(\cdot), T > t)$ 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 $(j = 1, 2)$ at time t does not further depend on the possibly unobserved CHD event time T. This fundamental assumption is called 'no unmeasured confounders for censoring' [10] and is equivalent to a sequential version of Rosenbaum and Rubin's strong ignorability assumption [11]. 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 missing at random (MAR) in the terminology of missing data analysis [3]. In practice, we would not expect this assumption to be precisely true, but given a rich collection of prognostic factors recorded in $\overline{V}_t$, it may well be approximately true.

Estimation of the probability of remaining in the study (Reason 1)

The IPCW approach is to artificially regard subjects as dependently censored at the first time a subject was censored by either loss to follow-up or refusal of 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 [7]. For the drop-outs due to loss to follow-up (Reason 1), the time-dependent Cox proportional hazards model for censoring is used for the right-hand side of Equation (1),

$$\lambda_{C_1}(t|R, \overline{V}_t, T > t) = \lambda_{0R}(t) \exp(\alpha_R \overline{V}_t), \quad (2)$$

where the treatment group-specific baseline hazard $\lambda_{0R}(t)$ and the treatment group-specific regression parameters α_R are assumed, because both the baseline hazard and covariate effects may depend on treatment group. For estimating the hazard of censoring (2) 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 due to the Reason 1 for subject i is provided by the following time-dependent extension of the Kaplan–Meier estimator,

$$\hat{K}_{i1}(t) = \prod_{u; X_u < t, \sigma_{u1}=1, R_u=R_i} \exp\left[-\hat{\lambda}_{0R}(X_u) \exp(\hat{\alpha}_R \overline{V}_{iX_u})\right], \quad (3)$$

where $\hat{\lambda}_{0R}(X_u) = \sigma_{u1} / \sum_{i=1}^{n} \exp(\hat{\alpha}_R \overline{V}_{iX_u}) Y_i(X_u) I \times (R_i = R)$ is the Breslow estimator of the baseline hazard function for censoring $j = 1$ 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_{u1} takes the value of one if the subject is censored for Reason 1, and zero otherwise. For any proposition A, $I(A)$ equals one if A is true and zero otherwise.

Estimation of the probability of remaining in the study (Reason 2)

For the drop-outs due to refusal of follow-up by patients (Reason 2), because the drop-out times are fixed at 5 years, the probability of drop-outs is modeled by the following logistic regression model,

$$\text{logit } \Pr(D_i = 1|R, \overline{V}_5, Z_i = 1) = \gamma_{0R} + \gamma_R \overline{V}_5, \quad (4)$$

where D_i takes the value of one if subject i refuses further follow-ups at 5 years, and zero otherwise, Z_i takes the value of one if subject i experiences the re-informed consent at 5 years, zero otherwise, and γ_{0R} and γ_R are the treatment group-specific regression parameters. Under the assumption of no unmeasured confounders for censoring (1) and the model (4), the conditional probability of being uncensored due to Reason 2 for subject i is estimated by

$$\hat{K}_{i2}(5) = 1 - \hat{P}r(D_i = 1|R, \overline{V}_5, Z_i = 1). \quad (5)$$

Estimation of the IPCW survival function

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 estimators of uncensored probability, $\hat{K}_{i1}(t)$ and $\hat{K}_{i2}(5)$, 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,

$$\hat{W}_i(t) = \begin{cases} \dfrac{1}{\hat{K}_{i1}(t)} & \text{for } t < 5 \\ \left(\dfrac{1}{\hat{K}_{i1}(t)}\right) \times \left(\dfrac{1}{\hat{K}_{i2}(5)}\right) & \text{for } t \geq 5 \text{ and } Z_i = 1. \\ \dfrac{1}{\hat{K}_{i1}(t)} & \text{for } t \geq 5 \text{ and } Z_i = 0 \end{cases}$$

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 $\overline{V}_t$. Under this assumption, the IPCW Kaplan–Meier estimator of the treatment group-specific survival of not having CHD events through time t is

$$\hat{S}_T(t|R) = \prod_{\{i;X_i<t\}} \left\{ 1 - \frac{\delta_i \hat{W}_i(X_i) I(R_i = R)}{\sum_{u=1}^{n} Y_u(X_i) \hat{W}_u(X_i) I(R_u = R)} \right\} \tag{6}$$

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 (6), 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 drop-outs, while the quantity, $\sum_{u=1}^{n} Y_u(X_i) \hat{W}_u(X_i) I \times (R_u = R)$, estimates the number of subjects who would have been alive and at risk at time X_i in the absence of drop-outs. Thus, the ratio estimates the hazard of CHD event at X_i in the absence of drop-outs; it follows that (6) estimates the probability $S_T(t|R)$ of surviving without failure (i.e., of remaining CHD-free) until time t in the absence of drop-outs. Under assumption (1) and the correct specification of weights, Robins [5] proves that under mild regularity conditions, the IPCW estimator (6) gives a consistent estimator of our target causal estimand $S_T(t|R)$. Inverse probability weighted estimators have been previously considered by Horvitz and Thompson [12] in the sample survey literature. Satten and Datta [13] give an elementary discussion of the IPCW estimators.

Comparison of the IPCW survival function

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

$$\lambda_T(t|R) = \lambda_0(t) \exp(\beta R)$$

where $\lambda_T(t|R)$ is the potential hazard of CHD events at time t in the 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,

$$U(\beta) = \sum_{i=1}^{n} \delta_i \hat{W}_i(X_i) \times \left\{ R_i - \frac{\sum_{u=1}^{n} Y_u(X_i) \hat{W}_u(X_i) R_u \exp(\beta R_u)}{\sum_{u=1}^{n} Y_u(X_i) \hat{W}_u(X_i) \exp(\beta R_u)} \right\} \tag{8}$$

Under the assumption (1) and the correct specification of weights, Robins [5] proves that under mild regularity conditions, the weighted estimating equations $U(\beta) = 0$ gives a consistent and asymptotically normal estimator of the parameter β, which can be interpreted as the treatment effect in the absence of drop-outs.

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 [14]. 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 [14,15].

Simulation studies

Settings of simulations

To evaluate the performance of the IPCW estimation method, we carried out simulation studies under three conditions: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). We simulated data from two treatment groups, coded as $R = 0$ (control treatment) or $R = 1$ (test treatment). About equal sample size of 500 for each group was randomly generated (total sample size was 1000). The simulations were based on 1000 replications, so that the estimated coverage probability of a true

95% confidence interval would have a simulation accuracy of ≈1.35%.

For each subject i ($i=1,\ldots,1000$), a time-dependent covariate L_{it} ($t=0,\ldots,4$) was generated via the following mixed effect model,

$$L_{it} = 10 - 0.10 \times I(R_i = 0) \times t - 1.70 \times I(R_i = 1) \times t + b_{0i} + b_{1i} \times t + \varepsilon_{it}.$$

The random effects b_{0i} and b_{1i} were generated from a bivariate normal distribution with means of zero and their variance of 3.0 and 2.5, respectively, with the correlation coefficient of 0.8. The random error ε_{it} was generated from a normal distribution with a mean of zero and a variance of 0.8. For each subject, L_{it} was supposed to be observed until just before the observed failure time.

The potential failure time T_i was generated from the following exponential distribution with hazard λ,

$$S(t) = \exp(-\lambda t),$$

where $\lambda = \exp(\alpha_0 + \alpha_1 R_i + \alpha_2 L_{i0})$, $(\alpha_1, \alpha_2) = (-0.5, 0.3)$, and $\alpha_0 = -6.0$ (MCAR and MAR cases), $\alpha_0 = -5.7$ (MNAR case). The drop-outs were assumed to occur at four time points ($t=1,\ldots,4$) and censoring at the end of follow-up ($t=5$) was considered independent censoring. For simplicity, only one type of drop-out was considered, where the drop-out indicator variable D_{it} ($D_{it}=1$ if drop-outs, $D_{it}=0$ if otherwise) was generated from the following conditional model,

$$\text{logit } \Pr(D_{it} = 1 | D_{i(t-1)} = 0, L_{i(t-1)}, t, T_i > t) = \beta_0 + \beta_1 t + \beta_2 L_{i(t-1)} + \beta_3 T_i, \quad (9)$$

where $t=1,\ldots,4$, $\beta_2=\beta_3=0$ corresponds to MCAR case (β_0 and β_1 were set to be -2.0 and -0.4, respectively), $\beta_2 \neq 0$, $\beta_3 = 0$ corresponds to MAR case (β_0, β_1, and β_2 were set to be -6.0, -0.4, and 0.4, respectively), and $\beta_3 \neq 0$ corresponds to MNAR case (β_0, β_1, β_2, and β_3 were set to be -3.5, -0.4, 0.4, and -0.3, respectively). In the above settings, although the potential failure time T_i is assumed to be directly dependent on group and baseline covariate L_{i0}, because of the high correlation between L_{i0} and L_{it}, the larger the values of L_{it}, the T_i is shorter and the more drop-out cases will be observed. The percentages of event, drop-outs, and censoring at the end of follow-up are roughly 10, 20, and 70% ($R=1$), 20, 20, and 60% ($R=0$), respectively. The observed failure time X_i was set to $X_i=t$ for drop-out cases at time t, $X_i=T_i$ for event cases whose potential failure time is not exceeding 5, and $X_i=5$ for censoring cases at the end of follow-up whose potential failure time is >5.

In each repetition of simulations, the proportional hazards model including a group variable R as a covariate was fitted to the observed failure time X_i and the estimate of the log(hazard ratio), $\hat{\theta}_s$ ($s=1,\ldots,1000$), was calculated. The following three proportional hazards models were fitted. The first one was the standard analysis ignoring the time-dependent covariate L_{it}, where all drop-out cases were assumed to be censored at their drop-out times (assumption of independent censoring). The second one was the adjusted analysis including the time-dependent covariate L_{it} as covariates under the assumption of independent censoring. The third one was the proposed IPCW analysis, where the weights were estimated by fitting the model (9) with $\beta_3=0$ to the observed data.

The result from the analysis for data that had been observed in the absence of drop-outs was regarded as a true value of the log(hazard ratio) in each repetition. The observed failure time that had been observed in the absence of drop-outs A_i was defined to be $A_i=T_i$ for event cases whose potential failure time is not exceeding 5, and $A_i=5$ for the censoring cases at the end of follow-up whose potential failure time is >5.

Results of simulations

Simulations were evaluated in terms of the percent relative bias, mean squared error (MSE), and coverage probability of nominal 95% large sample confidence intervals for the estimate of the log(hazard ratio) for group effect. The percent relative bias was computed as $(1/1000)\sum(\hat{\theta}_s - \bar{\theta})/\bar{\theta} \times 100$, where $\hat{\theta}_s$ is the estimate of $\bar{\theta}$ (average of true values for the log(hazard ratio)) from the sth simulated replication.

Table 3 shows the results. Under the MCAR setting, both estimates from the standard and the IPCW analysis were nearly unbiased and their coverage probabilities were close to the nominal level of 95%, while the adjusted estimate was largely biased with anticonservative coverage probability. Under the MAR setting, as expected based on the theory, the selection bias due to the time-dependent covariate L_{it} was adjusted by the IPCW analysis, while the estimate from the standard analysis underestimated the treatment effect because the subjects with larger values of L_{it} and shorter potential failure time tended to drop out. MSE from the IPCW analysis was slightly larger than that of the standard one. Under the MNAR setting, the IPCW analysis could not adjust the selection bias due to the violation of the assumption of no unmeasured confounders for censoring,

Table 3 The results of simulations for the estimate of treatment effect

	Estimation method	True value	Relative bias (%)	MSE	95% coverage
MCAR	Standard	−0.485	0.60	0.023	95.0
	Adjusted		−44.23	0.072	71.0
	IPCW		0.62	0.023	95.1
MAR	Standard	−0.483	−3.20	0.027	94.2
	Adjusted		−43.43	0.072	73.2
	IPCW		−0.57	0.032	94.9
MNAR	Standard	−0.479	−3.31	0.025	94.6
	Adjusted		−29.57	0.045	84.2
	IPCW		−2.46	0.026	95.3

MSE: mean squared error
Standard analysis is the proportional hazards model including only a group variable as a covariate under the assumption of independent censoring. Adjusted analysis is the same model except for including the time-dependent covariates.

although the bias was slightly smaller than that of the standard one.

Analysis of MEGA study data

Factors affecting each drop-out

To construct the IPCW estimators, it is necessary to estimate the subject-specific weight $\hat{W}_i(X_i)$ conditional on time-dependent prognostic factors for failure. We have to choose variables for modeling the censoring process so as to make assumption (1) plausible. As causal interpretation of estimates depends on the correctness of (1), making the censoring process ignorable is more important than fitting a parsimonious model. However, because of the large number of potential prognostic factors included in $\overline{V}_t$, it may be useful to reduce these to a relevant subset. In general, for a time-dependent prognostic factor to cause selection bias or confounding, it must be a prognostic factor for both failure and censoring.

To estimate the subject-specific weight $\hat{W}_i(X_i)$, we used five time-dependent factors as well as twelve baseline factors shown in Table 1. Among baseline factors, missing data were observed in the values of BMI (0.24%), current smoking (0.18%), and drinking (0.17%). The missing values of BMI were imputed by the mean value of 23.8 (kg/m^2). The later two factors were imputed by zero (no smoking and no drinking, respectively). Five time-dependent factors were four lipids (TC, TG, HDL-C, and LDL-C) and treatment actually received. For the missing data of lipid values (21.5%), the regression imputations were separately conducted, where 12 baseline factors, allocation group, and the last observed lipid value were included as covariates in each prediction model. For the missing data of treatment actually received (10.5%), the last observation carried forward method was used to impute the missing values.

To estimate the treatment group-specific hazards of censoring due to Reason 1 and the probability of drop-outs at 5 years due to Reason 2, five combinations of covariates $\overline{V}_t$ were used in both Models (2) and (4), accounting for the multicollinearity of covariates. First one included all five time-dependent factors and twelve baseline factors as covariates (Model 1). Second one excluded all TC values and time-dependent treatment group from Model 1 (Model 2). Third one excluded all baseline lipid values from Model 2 (Model 3). Fourth one excluded all TC values, time-dependent TG, HDL-C, and LDL-C from Model 1 (Model 4). Last one included only significant variables in the Model 2 (Model 5). The values of TG and HDL-C were included into the above five models by taking their logarithm. For the five time-dependent factors, the most recent recorded values were included as covariates in the prediction model.

Table 4 shows the effect estimates of each factor associated with two types of censorings. The results from Model 3 are presented, because the results from the other models were similar to those shown in Table 4, except previous non-use of pravastatin also predicted drop-outs from the study. For Reason 1, patients who have hypertension, diabetes mellitus, or hypercholesterolemia medication history tended to remain in the study. In the pravastatin group, the higher the values of TG or LDL-C during the study period, the more drop-out cases that were observed. For Reason 2, patients with hypertension or lower values of TG or HDL-C during the study period were likely to consent to the further follow-up at 5 years.

Table 4 Factors affecting each drop-out (results from Model 3)

	Loss to follow-up (Reason 1)				Refusal of follow-up by patients (Reason 2)			
	Diet		Diet + pravastatin		Diet		Diet + pravastatin	
Factors	HR	95% CI	HR	95% CI	OR	95% CI	OR	95% CI
Baseline								
Age (years)	1.01	0.99, 1.02	1.00	0.99, 1.02	0.99	0.97, 1.01	1.00	0.98, 1.02
Women	1.08	0.85, 1.37	1.00	0.80, 1.27	0.80	0.55, 1.18	1.28	0.86, 1.89
BMI (kg/m^2)	1.01	0.98, 1.04	0.98	0.95, 1.01	1.10	0.97, 1.06	0.98	0.94, 1.03
Current smoker	1.13	0.88, 1.44	1.21	0.94, 1.54	1.14	0.75, 1.72	1.23	0.81, 1.85
Current drinking	1.14	0.91, 1.45	1.03	0.83, 1.28	0.81	0.55, 1.17	1.17	0.80, 1.70
Medication history	0.84	0.66, 1.08	0.63	0.48, 0.81	0.87	0.58, 1.29	0.70	0.45, 1.08
Hypertension	0.82	0.68, 0.98	0.91	0.77, 1.07	0.79	0.60, 1.05	0.76	0.57, 1.01
Diabetes	1.02	0.82, 1.25	0.72	0.58, 0.90	0.83	0.60, 1.18	1.01	0.73, 1.42
Time-dependent								
TG (mg/dL)	1.11	0.92, 1.35	1.30	1.08, 1.57	1.56	1.08, 2.25	1.72	1.20, 2.48
HDL-C (mg/dL)	0.82	0.54, 1.26	1.11	0.74, 1.67	3.14	1.54, 6.41	1.76	0.84, 3.69
LDL-C (mg/dL)	1.00	0.99, 1.01	1.01	1.01, 1.01	1.00	1.00, 1.01	1.00	1.00, 1.01

HR: hazard ratio; OR; odds ratio; CI: confidence interval; medication history: hypercholesterolemia medication history

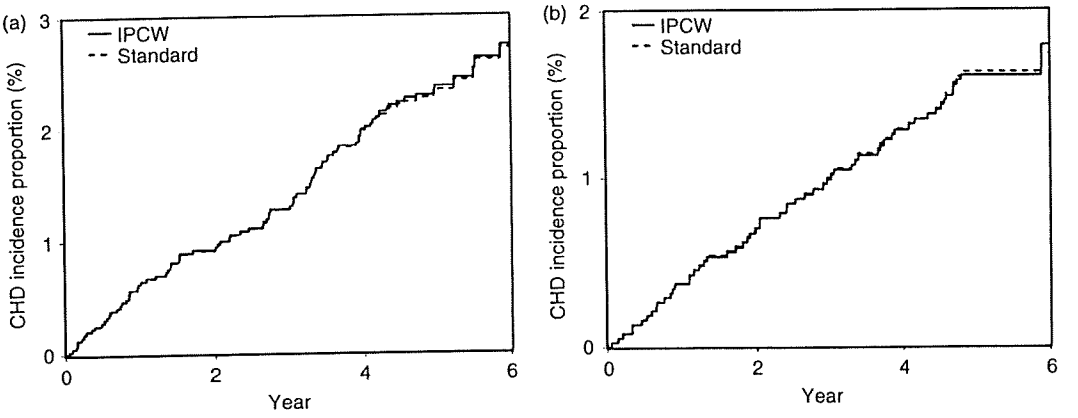

Figure 3 IPCW incidence proportion for CHD events in each treatment group. In each group, the solid line is the IPCW estimate and the dashed one is the standard estimate shown in Figure 2: (a) Diet group; (b) Diet + pravastatin group

Estimation of treatment effect adjusting for dependent censoring

Figure 3 shows the IPCW Kaplan–Meier curves for the CHD events in each treatment group. In each group, the solid line is the IPCW estimate whose weights were calculated from Model 1, and the dashed one is the standard estimate shown in Figure 2. In both treatment groups, the adjusted curves were almost the same as those obtained by assuming all drop-out cases as independent censoring. Table 5 shows the estimates of treatment effect under several models. 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 censorings were observed (not shown) as shown in Table 4. For both CHD and stroke events, a slightly larger treatment effect was observed by the IPCW analysis. The IPCW estimates did not change under different models for the estimation of weights.

Discussion

In this article, we developed a method for the estimation of treatment effect adjusting for

Table 5 Estimates of treatment effect for CHD and stroke

Method		CHD			Stroke		
		HR	95% CI	*p*-value	HR	95% CI	*p*-value
Standard		0.67	0.49, 0.91	0.010	0.83	0.57, 1.21	0.33
IPCW	Model 1	0.65	0.48, 0.89	0.007	0.81	0.56, 1.18	0.27
	Model 2	0.66	0.48, 0.90	0.008	0.81	0.56, 1.18	0.28
	Model 3	0.66	0.49, 0.90	0.009	0.81	0.56, 1.17	0.26
	Model 4	0.66	0.49, 0.91	0.009	0.81	0.56, 1.18	0.27
	Model 5	0.66	0.48, 0.90	0.008	0.82	0.57, 1.19	0.29

HR: hazard ratio; CI: confidence interval; standard method is the analysis assuming all types of censorings as independent.

dependent censoring using the IPCW approach. This proposed method is a straightforward extension of Robins and Finkelstein [7] method for settings with two or more reasons for censoring. In real clinical trials, there are several different types of reasons for censoring and it is likely that each process has its own reason, that is, the covariates history through each censoring. Our proposed approach was a relatively easy method for accounting for the differences in the reasons for censoring by estimating the weights separately in the framework of the IPCW methodology.

It is important to note that our results are based on the fundamental assumption (1) that the cause-specific hazard of censoring can be totally explained by the treatment group and the recorded history of the covariates. This assumption is a non-identifiable assumption and is not testable from the observed data. There is a possibility of a residual effect due to unmeasured prognostic factors for censoring. However, in the MEGA study, many clinically important prognostic factors were measured and all of them were used as covariates to predict the probability of remaining in the study. In addition to the five prediction models shown in Table 5, the analyses based on other prediction models, in which time-dependent covariates were entered in different ways, such as the difference from the baseline or the absolute past two values, were conducted, and the IPCW estimates were shown to be insensitive to the selection of the prediction models conditional on the measured covariates. Therefore, a departure from the assumption (1) will be small in our IPCW estimates.

In this article, we regarded both the institution-specific drop-out at 5 years and the end-of-study censoring as independent censoring. However, there may be a possibility of correlation between such censorings and the prognosis [16]. We also conducted the IPCW analyses, where all types of censorings were considered as potentially dependent ones. In this analysis, censoring due to institutional refusal at 5 years was separately modeled by the logistic regression model such as (4), and the end-of-study censoring was modeled by specifying the cause specific hazard functions, where the time-dependent Cox proportional hazards models such as (2) were separately fitted for each cause of censoring. The IPCW hazard ratios from Model 3 in combinations of covariates $\overline{V}_t$ were 0.66 (95% CI: 0.48–0.90) for CHD events and 0.81 (95% CI: 0.56–1.18) for stroke event. Therefore, our informal assumption that the end-of study censoring including institutional refusal at 5 years was considered as independent censoring seemed to be reasonable.

In the analyses of the MEGA study, factors affecting drop-outs were different for each reason for censoring as well as for the treatment group. However, no history of medication for hypertension, diabetes mellitus, or hypercholesterolemia was related to both censorings, that is, patients with a relatively better medical condition tended to drop out. This suggested that troublesomeness or weak motivation for participating in the study might cause the drop-outs, because the MEGA study was a primary prevention study. Furthermore, the fact that patients with high HDL-C or not in the pravastatin group tended to drop-out may also explain the above possibility for the censoring process in the MEGA study. This censoring process was different from that observed in usual clinical trials where the occurrence of adverse events or deteriorating health condition are the primary reasons for study drop-out. On the other hand, for the time-dependent covariates, patients with higher values of TG tended to drop out, suggesting that non-compliance with the appropriate diet instructions or inadequate diet control during the study period were related to the censoring processes.

Compared with the standard analysis, a slightly larger treatment effect was observed by the IPCW analysis, but the difference was minimal.

In general, selection bias is a function of both the magnitude of censoring rate and how different the censored subjects are from uncensored ones in terms of prognosis. In the MEGA study, although the censoring proportions due to patient refusal of follow-up (Reason 2) were relatively small and there were no differences between treatment groups, those due to loss to follow-up (Reason 1) were relatively large and the differences were observed between treatment groups (Figure 1(a)). In the latter category of censoring, about half of the reasons for loss to follow-up were the withdrawal of informed consent (51.8% in diet group, 64.3% in pravastatin group). As shown in Table 4 and discussed previously, the censoring process observed in the MEGA study seemed to be unrelated to the occurrences of outcomes of interest. Therefore, the lack of effect of weighting in our data could be due to the fact that the weights are not highly related to the probabilities of disease outcome and thus would not have an appreciable effect in altering the point estimates. This result indicates that drop-outs observed in the MEGA study did not cause a severe selection bias attributable to the measured covariates and the standard analysis results [2] were robust for the drop-outs.

However, as shown in the results of the simulations, the estimate from the standard analysis is biased under the MAR setting where the drop-out process is dependent on the covariate histories. In many clinical trials, because we cannot safely say that the dependent censorings have not occurred, it is important to conduct the analysis accounting for the dependent censorings as well as the standard one and to compare their results. When their results differ remarkably, the reasons for drop-outs were examined in detail and the effects on the final conclusion in the clinical trial concerned should be discussed.

Acknowledgments

MEGA Study is supported by Sankyo Co. LTD. We also thank the referee, the associate editor, and the editor for their comments, which led to a much improved version of the article.

References

1. 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. *Circ J* 2004; 68: 860–67.
2. 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: 1155–63.
3. Rubin DB. Inference and missing data. *Biometrika* 1976; 63: 581–92.
4. Kalbfleish JD, Prentice RL. *The Statistical Analysis of Failure Time Data* (2nd ed.) John Wiley, New York, 2002.
5. 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.
6. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J A Stat Assoc* 1995; 90: 106–21.
7. 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: 779–88.
8. Scharfstein DO, Robins JM. Estimation of the failure time distribution in the presence of informative censoring. *Biometrika* 2002; 89: 617–34.
9. Joffe MM, Hoover DR, Jacobson LP *et al.* Estimating the effect of zidovudine on kaposi's sarcoma from observational data using a rank preserving structural failure-time model. *Stat Med* 1998; 17: 1073–102.
10. Robins JM. Causal inference from complex longitudinal data. In Berkane M. Ed). *Latent variable modeling and application to causality* Springer-Verlag, New York, 69–117, 1997.
11. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70: 41–55.
12. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J A Stat Assoc* 1952; 47: 663–85.
13. Satten GA, Datta S. The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician* 2001; 55: 207–10.
14. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *J A Stat Assoc* 1989; 84: 1074–78.
15. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11: 550–60.
16. Zhang J, Heitjan DF. Nonignorable censoring in randomized clinical trials. *Clin Trials* 2005; 2: 488–96.

STATISTICS IN MEDICINE
Statist. Med. (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3065

Estimation of treatment effect adjusting for treatment changes using the intensity score method: Application to a large primary prevention study for coronary events (MEGA study)

Yukari Tanaka*,†, Yutaka Matsuyama and Yasuo Ohashi for the MEGA Study Group

Department of Biostatistics/Epidemiology and Preventive Health Sciences, School of Health Sciences and Nursing, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

SUMMARY

The MEGA study was a prospective, randomized, open-labeled, blinded-endpoints study conducted in Japan to evaluate the primary preventive effect of pravastatin against coronary heart disease (CHD), in which 8214 subjects were randomized to diet or diet plus pravastatin. The intention-to-treat (ITT) analysis showed that pravastatin reduced the incidence of CHD (hazard ratio = 0.67; 95 per cent confidence interval (CI): 0.49–0.91) and of stroke events, which was the secondary endpoint in the MEGA study (hazard ratio = 0.83; 95 per cent CI: 0.57–1.21). Owing to considerable treatment changes, it is also of interest to estimate the causal effect of treatment that would have been observed had all patients complied with the treatment to which they were assigned. In this paper, we present an intensity score method developed for clinical trials with time-to-event outcomes that correct for treatment changes during follow-up. The proposed method can be easily extended to the estimation of time-dependent treatment effects, where the technique of g-estimation has been difficult to apply in practice. We compared the performances of the proposed method with other methods (as-treated, ITT, and g-estimation analysis) through simulation studies, which showed that the intensity score estimator was unbiased and more efficient. Applying the proposed method to the MEGA study data, several prognostic factors were associated with the process of treatment changes, and after adjusting for these treatment changes, larger treatment effects for pravastatin were observed for both CHD and stroke events. The proposed method provides a valuable and flexible approach for estimating treatment effect adjusting for non-random non-compliance. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: non-compliance; time-dependent confounding; causal inference; failure time data; propensity score; structural nested mean model

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

Contract/grant sponsor: Grant-in-Aid for Scientific Research; contract/grant number: 16200022
Contract/grant sponsor: Japanese Ministry of Health, Labor and Welfare
Contract/grant sponsor: Sankyo Co Ltd, Tokyo

Received 8 December 2006
Accepted 10 August 2007

1. INTRODUCTION

In a typical clinical trial, patients are randomized to one of the treatment groups and each patient is expected to receive that treatment throughout the follow-up to assess the effect of the treatment on some outcome. However, most clinical trials are not ideal; hence, patients often fail to adhere to their assigned treatment and switch to another trial treatment. Such non-compliance with assigned treatment is a common feature of clinical trials. Recently there has been much interest in methods for analyzing randomized clinical trials of treatments to which the subject are not compliant [1–3].

One approach for analyzing data for non-compliance is the as-treated (AT) analysis, which compares outcomes based on the treatment that patients actually received. When non-compliance is completely at random, that is, independent of both (observed and unobserved) baseline and time-dependent factors, the AT analysis can give a valid test for the null hypothesis of no treatment effect and can also give an unbiased estimate of treatment effect. In most clinical trials, however, patients who comply with their assigned treatment are not comparable with those who do not with respect to some important prognostic factors. In this case, both the decision to comply and the outcome may well depend on underlying possibly unmeasured health status. Thus, when non-compliance is non-random, the AT analysis will not be valid even under the null hypothesis because of the comparison of selected groups [3, 4].

The more commonly used analytic approach is an intention-to-treat (ITT) analysis, which compares outcomes based on the treatment groups randomized by design regardless of whether the patients complied with their assigned treatment. Because the comparability of the treatment groups is guaranteed by randomization, the null hypothesis of no treatment effect for all patients (sharp causal null hypothesis) is preserved in the ITT analysis. That is, successful randomization insures that the ITT comparison provides a valid test for the sharp causal null hypothesis of no treatment effect even in the presence of non-random non-compliance. Moreover, p-values have a randomization interpretation when design-based (randomization-based) analyses are used [5]. Furthermore, the ITT estimate would correspond to the overall treatment effect that would be realized if the treatment were actually adopted and practiced in the community, provided the rate of non-compliance and the factors influencing non-compliance that are observed in the trial are identical to those that would occur in the community. A point against the ITT analysis is that the ITT parameter does not measure the true biological effect of treatment, but rather a mixture of the effect on the compliers with the absence of effect on the non-compliers, because the ITT estimate is the average effect of treatment assignment. Hence, the ITT analysis gives estimates that are biased toward the null when treatment crossover is present, and the ITT measure of treatment effect will diminish as non-compliance increases. Moreover, the rate of non-compliance in the community, once the treatment is adopted, may not be the same as the rate in the original clinical trial.

Therefore, in the analysis of non-compliance data, it is important to estimate the causal effect of treatment, that is, the effect that would be realized if all patients complied with the treatment to which they were assigned. Robins [6–8] has proposed a structural nested mean model (SNMM) to estimate such causal effect in the presence of non-random non-compliance. Under the assumption that non-compliance at each time is at random, given the observed histories that influence a patient's decision to comply, that is, the assumption of no unmeasured confounders, the causal parameter in a SNMM can be estimated by the technique of g-estimation.

Recently, Brumback *et al.* [9] proposed the intensity score approach for the analysis of time-varying treatments in the presence of time-dependent confounding. They provided conditions

DOI: 10.1002/sim

under which the intensity score approach consistently estimates a treatment effect in a SNMM. The intensity score is cumulative differences over time between treatment actually received and treatment predicted by prior observed medical history. The SNMM treatment effect can be obtained by regressing outcomes on the intensity score. Thus, the intensity score approach can provide an easy implementation of g-estimation for the analysis of non-random non-compliance. Since the intensity score approach was originally proposed for continuous outcomes, we extend its use to time-to-event outcomes with censoring. This extension is useful, because censoring due to end of scheduled follow-up requires special care when using g-estimation based on the structural accelerated failure time (SAFT) model [10–16], while the intensity score approach can treat the censoring within the framework of standard regression models. Furthermore, the intensity score approach has the advantage of providing estimates of parameters in a SNMM that allows the treatment effects to vary across time, while it has been difficult to apply such a model in practice using the technique of g-estimation [9].

This article is organized as follows. In Section 2, we describe the motivating study from a large randomized primary prevention study for coronary events, the Management of Elevated Cholesterol in the Primary Prevention Group of Adult Japanese (MEGA) study [17, 18]. In Section 3, we develop the intensity score approach for event times. In Section 4, simulation studies are conducted to compare the performances of the proposed intensity score method with those of the AT, ITT, and g-estimation (semi-parametric randomization-based) analysis [10, 12]. Section 5 presents the analysis results of the MEGA study data. Finally, Section 6 provides some discussions.

2. THE MEGA STUDY

We will briefly describe the motivating study and the data (MEGA study). Full details on the design, conduct, and main clinical results have been reported [17, 18]. The MEGA study is a randomized controlled trial conducted in Japan to evaluate the primary preventive effect of a statin against coronary heart disease (CHD) in daily clinical practice. In this prospective, randomized, open-labeled, blinded-endpoints design study, men and postmenopausal women aged 40–70 years with hypercholesterolemia (total cholesterol (TC) level: 220–270 (mg/dL)) and no history of CHD or stroke were randomized to diet (diet group) or diet plus pravastatin 10–20 mg daily (pravastatin group).

Between February 1994 and March 1999, a total of 15 210 persons visiting outpatient clinics were registered throughout Japan. Of the 15 210 subjects who met the inclusion criteria regardless of their TC levels and who provided signed informed consent, 8214 who met the TC criterion were randomized to either diet or diet plus pravastatin treatment using the permuted block method with stratification according to gender, age, and medical institution. After the exclusion of 382 patients (94 withdrew consent, 224 exclusion criteria violation, and 64 no recorded data after randomization), the remaining 7832 patients were analyzed (3966 diet group; 3866 pravastatin group).

Table I shows the baseline characteristics of the analyzed patients. There was no clinical difference between the two groups in baseline characteristics. Women accounted for 68.4 per cent (5356 patients) of the study population. Mean body mass index (BMI) was 23.8 (kg/m^2). Mean TC, low-density lipoprotein cholesterol (LDL-C), and high-density lipoprotein cholesterol (HDL-C) levels were 242.6, 156.6, and 57.5 (mg/dL), respectively. Median triglyceride (TG) level was 127.5 (mg/dL). Of the study patients, 41.8 and 20.8 per cent had hypertension and diabetes mellitus based on physician diagnosis, respectively.

DOI: 10.1002/sim

Table I. Baseline characteristics of analyzed 7832 patients.

Characteristics	Diet group ($N=3966$)		Diet+pravastatin group ($N=3866$)	
Age (years), mean (SD)	58.4	(7.2)	58.2	(7.3)
Women, no. (per cent)	2718	(68.5)	2638	(68.2)
BMI (kg/m^2), mean (SD)	23.8	(3.0)	23.8	(3.1)
Current smoker, no. (per cent)	572	(14.4)	612	(15.8)
Current drinking, no. (per cent)	1183	(29.8)	1180	(30.5)
Hypercholesterolemia medication history, no. (per cent)	621	(15.7)	586	(15.2)
Hypertension, no. (per cent)	1664	(42.0)	1613	(41.7)
Diabetes, no. (per cent)	828	(20.9)	804	(20.8)
TC (mg/dL), mean (SD)	242.6	(12.1)	242.6	(12.0)
TG (mg/dL), median (inter-quartile range)	127.5	(95.0–179.0)	127.4	(95.7–176.5)
HDL-C (mg/dL), mean (SD)	57.5	(15.1)	57.5	(14.8)
LDL-C (mg/dL), mean (SD)	156.5	(17.3)	156.7	(17.6)

SD, standard deviation; BMI, body mass index; TC, total cholesterol; TG, triglyceride; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

After randomization, patients were followed at months 1, 3, and 6 and thereafter every 6 months. At each visit, data on treatment compliance, use of concomitant drugs, onset of events, occurrence of adverse events, and laboratory tests including serum lipids were collected by the investigators. Additionally, an ECG (electrocardiogram) was obtained and evaluated annually. The follow-up period was initially scheduled for 5 years; however, on the basis of the recommendation of the Data and Safety Monitoring Committee, the study was continued for an additional 5 years to increase the number of events, and thus, patients who provided written consent at 5 years to continue the study were followed until the end of March 2004 [17, 18].

The primary endpoint was the first occurrence of CHD, comprised of fatal and non-fatal myocardial infarction, angina, cardiac and sudden death, and a coronary revascularization procedure. One of the secondary endpoints was the first occurrence of stroke events. All endpoints were reviewed strictly by the blinded Endpoint Committee and additional information obtained from the physician as needed [17]. A total of 7832 patients were followed by 2658 physicians in 1320 hospitals. The follow-up period was 41 195 person-years (mean follow-up period 5.3 years). CHD events occurred in 101 of 3966 patients in the diet group (2.55 per cent) and 66 of 3866 patients in the pravastatin group (1.71 per cent). Figure 1 shows the Kaplan–Meier curves for CHD events. The ITT analysis indicated that the incidence of CHD was significantly lower by 33 per cent in the pravastatin group than in the diet group (The ITT hazard ratio$=0.67$; 95 per cent confidence interval (CI): 0.49–0.91; $p=0.01$ for the log-rank test) [18].

However, many patients changed to the other trial treatment frequently during the study period (treatment crossover). This was because the protocol in the MEGA study stated that patients in the diet group could be switched to pravastatin treatment when a reduction of TC level was not observed, while patients in the pravastatin group could discontinue pravastatin treatment when the reduction of TC level was observed. The treatment decisions for changing the treatment or increasing the dose of paravastatin were determined by each treating physician. Patients who changed to another trial treatment even once in the first 5 years were 19.9 per cent ($n=790$) in the diet group and 53.4 per cent ($n=2064$) in the pravastatin group. These numbers for the whole

DOI: 10.1002/sim

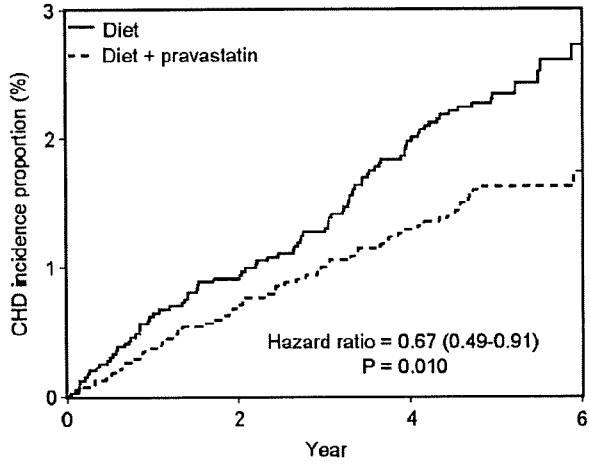

Figure 1. Incidence proportion for CHD events.

10 years were 21.3 per cent ($n=844$) and 63.1 per cent ($n=2441$), respectively. The effect of patients from one treatment to the other is to make the treatment profiles of the two randomized groups more similar than they otherwise would have been, and therefore to move the ITT hazard ratio toward the null.

3. INTENSITY SCORE METHOD

3.1. The multiplicative structural nested mean model

We consider a randomized clinical trial in which two groups (test and control treatment) are compared with respect to time-to-event outcomes and each patient i ($i=1,\ldots,N$) receives one of the treatments at the start of each time t ($t=0,\ldots,M-1$; time zero is the randomization time and the start of the first treatment). However, some patients fail to comply with their assigned treatment and cross over to the other treatment at each time t.

Suppose we have repeated measures on treatment $S_i(t)$ ($S_i(t)=1$ if test treatment, $S_i(t)=0$ if control treatment) and covariates $L_i(t)$ at time t. Let $H_i(t)$ be the observed history of treatment and the covariates prior to treatment at time t, i.e. $H_i(t)=(L_i(0), S_i(0),\ldots, L_i(t-1), S_i(t-1), L_i(t))$, with $H_i(0)=(L_i(0))$. Let $T_i(\overline{S}_i(t), 0)$ denote the potential event times in response to the hypothetical treatments $(S_i(0),\ldots, S_i(t), S_i(t+1)=0,\ldots, S_i(M-1)=0)$. That is, $T_i(\overline{S}_i(t), 0)$ represents the event time we would have observed if, possibly contrary to fact, the patient had his/her actual treatment history up to time t but was then switched to control treatment at time $t+1$ and remained at that treatment until the event occurred. Our notation for the potential outcomes implicitly assumes Rubin's stable unit treatment value assumption, which implies that potential outcomes of patient i do not depend on the treatment received by any other patient [19]. We will also assume that the potential outcomes satisfy the consistency assumption [7] that serves to link the potential outcomes with the observed outcomes T_i. This assumption states that $T_i=T_i(\overline{S}_i(t), 0)$ for all t when actually $S_i(t+1)=\cdots=S_i(M-1)=0$ occurred.

DOI: 10.1002/sim