[16, 17]. A recent prospective randomized controlled trial showed that DGE occurred in 50% of patients in whom the retrocolic route was used but in only 5% in whom the antecolic route was used [18]. These data suggest that the antecolic route is better. However, the 50% incidence of DGE associated with the retrocolic route seems high. We have shown that vertical retrocolic duodenojejunostomy, by which the stomach and duodenum are brought down through the left side of the transverse mesocolon in a straight, vertical manner, reduces the incidence of DGE [19].

Thus, a prospective randomized controlled trial was conducted to compare the incidence of clinical DGE and gastric emptying variables assessed by the ^{13}C-acetate breath test between patients who underwent antecolic duodenojejunostomy and those who underwent vertical retrocolic duodenojejunostomy. The aim of the study was to establish the superiority of the vertical retrocolic route with respect to gastric emptying after PpPD.

Patients and methods

Of 50 patients underwent pancreaticoduodenectomy at Miyazaki University Hospital between March 2005 and July 2007, 46 patients were scheduled to undergo PpPD. Patients were recruited into the study before surgery, on the basis of whether PpPD was anticipated and informed consent was obtained. Specific exclusion criteria included tumor infiltration into the duodenal bulb or presence of lymph node metastasis of the prepylorus ($n = 3$), failure to provide informed consent including the ^{13}C-acetate breath test ($n = 4$) were then excluded. Thus, 35 patients who underwent PpPD and consented to the protocol were enrolled in the study.

This prospective randomized controlled trial was approved by the ethical committee of our university hospital and informed consent was obtained from all patients. The randomization protocol involved assignment of patients to one of two reconstruction methods, the antecolic route and the vertical retrocolic route. Randomization took place during surgery before reconstruction. Gastric emptying was evaluated by means of the ^{13}C-acetate breath test just before surgery and on postoperative day (POD) 30.

Operative technique

The area resected during PpPD included the gallbladder, common hepatic duct, head of the pancreas, duodenum (except for the first portion), and 10 cm of the proximal jejunum. A few arcades of the right gastric artery and right gastroepiploic artery to the stomach were divided along the wall of the antrum (approximately 2–3 cm from the pyloric ring) for dissection of the peripyloric lymph nodes. The duodenum was freed from the surrounding tissue and transected approximately 4–5 cm distal to the pyloric ring. The lymph nodes in the hepatoduodenal ligament, the para-aortic lymph nodes, and those along the common hepatic artery and the right side of the superior mesenteric artery were dissected. The right gastric artery was divided at its origin in all patients. The left gastric artery and vein were carefully preserved. The lesser omentum close to the liver was dissected to allow free movement of the stomach. The vagal nerve, with the exception of the hepatic and pyloric branches, was preserved. These procedures allowed the stomach and the duodenum to be mobilized to the left in a straight, vertical manner.

As the first step in reconstruction, the proximal jejunum was brought through the right side of the transverse mesocolon by the retrocolic route. An end-to-side pancreaticojejunostomy was performed with duct-to-mucosal anastomosis. The pancreatic duct was anastomosed to the whole layer of the small opening in the jejunum to approximate the duct to the jejunal mucosa with the use of eight interrupted 5-0 PDS-II sutures (polydioxanone, Johnson & Johnson Co.), regardless of the size of the pancreatic duct. A 5-Fr polyethylene pancreatic drainage tube with a small knob (Sumitomo Bakelite Co., Japan) was placed in the pancreatic duct and exteriorized through the jejunal limb. The cut surface of the pancreas was then anastomosed to the jejunal seromuscular layer, and the end-to-side pancreaticojejunostomy was completed. A one-layer end-to-side hepaticojejunostomy with interrupted 5-0 PDS sutures was then performed 5–10 cm distal to pancreaticojejunostomy.

The final step was randomized to either to the antecolic route or vertical retrocolic route. For vertical retrocolic duodenojejunostomy, the left side of the transverse mesocolon (left side of the middle colic vessels) was opened, and the duodenum was brought down together with the gastric antrum in a straight, vertical manner. A retrocolic end-to-side duodenojejunostomy was performed at the caudal side of the transverse mesocolon and the antrum was fixed to the transverse mesocolon with a few 4-0 silk sutures. For antecolic duodenojejunostomy, the stomach was brought down antecolically. Braun anastomosis was added in both groups. Finally, the opening of the old ligament of Treitz and the jejunum brought up for pancreaticojejunostomy and hepaticojejunostomy were fixed to the mesocolon, and two or three closed drains were placed around the pancreatic and biliary anastomosis. All patients were given prophylactic antibiotics and H2 blocker postoperatively; none were given prokinetic drugs such as erythromycin.

Data collection and study endpoints

Clinicopathological data were collected prospectively for all patients. Data included postoperative mortality and morbidity, including pancreatic fistula, intraabdominal bleeding, pancreaticojejunostomy or hepaticojejunostomy leakage, intraabdominal abscess, and wound infection. Pancreatic fistula was defined when an amylase level in the fluid from the closed drains was >10,000 IU/l.

The first endpoint was clinical DGE defined as (1) the need for nasogastric tube decompression for more than 10 days (DGE 10), (2) the need for reinsertion of the nasogastric tube, or (3) an inability to take in an appropriate amount solid food orally by POD 14 (DGE 14), as described elsewhere [18].

The secondary endpoint was recovery of gastric emptying as assessed by ^{13}C-acetate breath test [20]. For at least 4 days before this test, all drugs, including H_2 blocker, were withdrawn. All patients ingested a liquid meal (200 Kcal/200 ml, RACOL, Ohtsuka Pharmaceutical Co., Tokyo, Japan) labeled with 100 mg sodium ^{13}C-acetate (Cambridge Isotope Laboratories, Inc., Andover, MA, USA) in the morning after an overnight fast before surgery and on POD 30. Breath samples were collected in the collection bag (1.3 l) before and after ingestion of the test meal, i.e., before and at 5, 10, 15, 20, 30, 40, 50, 60, 75, 90, 105, 120, and 180 min after ingestion of the ^{13}C-acetate. The recovery of ^{13}C in the breath samples was analyzed by isotope-selective infrared spectrometry (UBiT IR 300, Otsuka Electronics Co., Ltd, Osaka, Japan). The time when $^{13}CO_2$ reached maximum excretion (Tmax) and half-emptying time (T1/2) were calculated by using analysis software (Microsoft Office Excel, Microsoft Japan, Tokyo, Japan).

Statistical analysis

All values are expressed as mean ± SD. Differences between groups were examined for statistical significance by chi-square test, unpaired or paired Student's t-test, Wilcoxon signed-rank test, or Mann–Whitney U test. Statistical analysis was performed by the statistician who was blind to the study group.

Results

Patient characteristics

Clinical characteristics of the enrolled patients are shown in Table 1. There were no statistical differences between the two groups in age, sex ratio, type of disease, percentage of patients with malignant disease, preoperative laboratory data including *N*-benzoyl-L-tyrosyl-*p*-aminobenzoic acid (BT-PABA) test value, percentage of patients with diabetes mellitus, HbA1c, operation time, or length of the remaining duodenum.

Table 1 Patient characteristics

	Duodenojejunostomy reconstruction route		
	Antecolic $n = 17$	Vertical retrocolic $n = 18$	*P* value
Age (years)	69.7 ± 11.0	66.9 ± 12.9	0.50
Male/female ratio	11/6	9/9	0.38
Hemoglobin (g/dl)	11.8 ± 1.3	12.3 ± 1.5	0.30
Serum albumin (g/dl)	3.67 ± 0.31	3.71 ± 0.46	0.76
Total cholesterol (mg/dl)	171.6 ± 37.4	179.4 ± 38.4	0.55
Diabetes mellitus (+/−)	5/12	2/16	0.23
BT-PABA test (%)	52.3 ± 18.2	49.2 ± 16.1	0.60
HbA1c (%)	5.7 ± 1.6	5.5 ± 0.9	0.67
Soft pancreas	9	10	0.88
Operation time (min)	602.6 ± 93.5	581.7 ± 76.5	0.48
Blood loss (ml)	1619.4 ± 914.9	1535.0 ± 877.7	0.78
Residual duodenum (cm)	3.7 ± 0.7	3.8 ± 0.5	0.54
Division of right gastric artery	17	18	0.54
Final diagnosis			
Benign/malignant disease	5/12	2/16	0.23
Bile duct cancer	8	6	
Pancreatic cancer	4	6	
Ampullary cancer	0	2	
Duodenal cancer	0	1	
IPMN	2	1	
Chronic pancreatitis	2	2	
Benign bile duct tumor	1	0	

Values are mean ± SD or number of patients

IPMN intraductal papillary mucinous neoplasm, *BT-PABA* *N*-benzoyl-L-tyrosyl-*p*-aminobenzoic acid

Postoperative complications

As shown in Table 2, postoperative morbidity was observed in 9 of 17 patients (53%) in the antecolic group and 6 of 18 patients (33%) in the vertical retrocolic group. Intra-abdominal bleeding associated with pancreatic fistula and/or intra-abdominal abscess was observed in one patient in each group, and both patients were treated successfully by interventional transarterial embolization. Intra-abdominal abscess was the main complication and were treated successfully by drainage. No operative death or hospital death was observed.

Table 2 Postoperative outcomes

	Duodenojejunostomy reconstruction route		
	Antecolic $n = 17$	Vertical retrocolic $n = 18$	P value
Postoperative morbidity	9	6	0.24
Major P-J leakage	0	0	
Pancreatic fistula	1	1	0.97
H-J leakage	0	0	
Intra-abdominal bleeding	1	1	0.97
Intra-abdominal abscess	6	5	0.63
Wound infection	3	3	0.94
Respiratory dysfunction	0	0	
G-I bleeding	1	0	0.49
D-J leakage	0	0	
Mortality	0	0	
NG tube removed (POD)	1.2 ± 0.4	1.1 ± 0.3	0.59
DGE10	0	0	
Reinsertion of NG tube	0	0	
Liquid meal begun (POD)	5.4 ± 2.7	5.7 ± 2.4	0.72
Solid foods begun (POD)	8.4 ± 3.0	10.2 ± 5.1	0.21
DGE14	1	4	0.34
Postoperative stay (days)	40.8 ± 12.3	39.4 ± 11.1	0.74

P-J pancreaticojejunostomy, *H-J* hepaticojejunostomy, *G-I* gastrointestinal, *D-J* duodenojejunostomy, *NG* nasogastric

Clinical DGE

DGE clinically defined as DGE10 or DGE14 and the length of postoperative hospital stay are shown in Table 2. The nasogastric tube was removed on POD 1.2 ± 0.4 in the antecolic group on POD 1.1 ± 0.3 in the vertical retrocolic group. No patient needed a nasogastric tube for more than 10 days (DGE10), and reinsertion of a nasogastric tube was not necessary in any patient. The number of days to the start of liquid diet was similar between the two groups (5.4 days in the antecolic group and 5.7 days in the vertical retrocolic group). With respect to DGE14, one patient in the antecolic group and four in the vertical retrocolic group failed unlimited solid food oral intake by POD 14. Thus, the incidence of DGE defined as DGE14 was 6% (1 of 17 patients) in the antecolic group and 22% (4 of 18 patients) in the vertical retrocolic group. Although the rate was higher in the vertical retrocolic group, the difference did not reach statistical significance ($P = 0.34$). The overall incidence of DGE after PpPD was 14% (5 of 35 patients).

^{13}C-acetate gastric emptying test

Tmax did not differ between the vertical retrocolic group and the antecolic group before or on POD 30 ($P = 0.56$ before surgery and $P = 0.31$ on POD 30). Similarly, T1/2 did not differ between the two groups ($P = 0.99$ before surgery, $P = 0.26$ on POD 30) (Table 3). Neither reconstruction route had a significant effect on gastric emptying on POD 30 after PpPD.

Table 3 ^{13}C-Acetate gastric emptying test results

	Duodenojejunostomy reconstruction route		
	Antecolic $n = 17$	Vertical retrocolic $n = 18$	P value
Before surgery			
Tmax (h)	1.11 ± 0.25	1.08 ± 0.29	0.56
T1/2 (h)	1.78 ± 0.31	1.92 ± 0.81	0.99
After surgery (POD 30)			
Tmax (h)	1.54 ± 1.22	2.12 ± 2.14	0.31
T1/2 (h)	3.63 ± 3.15	6.21 ± 8.62	0.26

Tmax the time when $13CO_2$ reached maximum excretion, *T1/2* half emptying time

The ^{13}C-acetate gastric emptying test values before and on POD 30 were compared in each group. In the vertical retrocolic group, Tmax was significantly prolonged on POD 30 compared to that before surgery (2.12 ± 2.14 h versus 1.08 ± 0.29 h, $P < 0.02$), whereas no significant difference was found in the antecolic group (1.54 ± 1.22 h versus 1.11 ± 0.25 h, $P = 0.29$). However, T1/2 was significantly longer in each group on POD 30 compared to the corresponding preoperative value ($P = 0.0023$ in the antecolic group, $P = 0.0002$ in the vertical retrocolic group (Table 3). Gastric emptying was not completely restored to the preoperative level in either group by POD 30. Mean Tmax on POD 30 in the antecolic group was increased 1.39-fold, and that in the vertical retrocolic group was increased 1.96-fold. Similarly, T1/2 was increased 2.04-fold in the antecolic group and 3.23-fold in the vertical retrocolic group. Greater increases in Tmax and T1/2 were observed in the vertical retrocolic group than in the antecolic group.

Tmax before and after surgery in each patient is shown in Fig. 1. Individual changes in Tmax before and after surgery were similar to individual changes in T1/2. A greater than twofold increase in Tmax was observed in 3 (18%) of 17 patients in the antecolic group, and in 4 (22%) of 18 patients in the vertical retrocolic group. Tmax of all patients before surgery ($n = 35$) was 1.09 ± 0.26 h, ranging from 0.7 to 1.5 h. Tmax greater than 1.5 h on POD 30 was found in four patients in each group. Tmax on POD 30 remained similar to the preoperative level in most patients (approximately 80%) in both groups.

Discussion

The present study showed that the incidence of clinical DGE was lower with the antecolic route than with the vertical

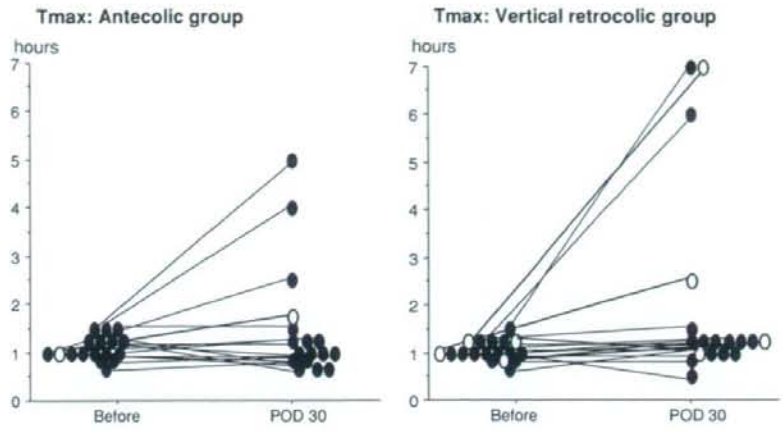

Fig. 1 Changes in Tmax of individual patients (before surgery versus POD 30). *Open circles* represent patients who were not able to tolerate appropriate solid food by POD 14 (DGE14)

retrocolic route, but the difference was not significant (6% with the antecolic route versus 22% with the vertical retrocolic route, $P = 0.34$). Moreover, gastric emptying (Tmax, T1/2) as assessed by the ^{13}C-acetate breath test did not differ significantly between the antecolic route and the vertical retrocolic route before or on day 30 after PpPD. T1/2 was significantly prolonged in both groups after PpPD, indicating that gastric emptying remained impaired on POD 30, regardless of the reconstruction route. The degree of impairment was greater in patients in whom vertical retrocolic reconstruction was performed. An analysis of individual patients revealed that on POD 30, gastric emptying was similar to the preoperative level in approximately 80% of patients, regardless of the reconstruction route.

Since Traverso and Longmire [1] first reported PpPD in 1978, the procedure has been accepted as a standard procedure for periampullary diseases. This is because it yields better quality of life, nutritional status, and weight gain without any difference in postoperative survival than the Whipple procedure [1–4, 21]. The postoperative mortality rate has fallen recently, but complications associated with pancreaticoduodenectomy remain, the most troublesome of which are pancreatic fistula, intra-abdominal infection, intra-abdominal bleeding, wound infection, and DGE. DGE was first reported by Warshaw and Torchiana [5]. Postoperative DGE decreases patient comfort, increases the risk of aspiration pneumonia, prolongs hospital stay, and increases medical costs.

DGE is considered a specific complication of PpPD, because it is specifically attributed to pylorus-sparing resection of the pancreatic head [5–7, 10]. Several underlying mechanisms have been proposed: (1) gastric atony or gastroparesis caused by vagotomy, resection of the duodenal pacemaker, or disruption of the gastroduodenal neural connections [11], (2) local ischemic injury of the antrum and pylorus [7], (3) gastric atony in response to a reduced circulating levels of motilin [12], (4) torsion or angulation of the reconstructed alimentary tract [7], (5) gastric dysrhythmia or gastroparesis secondary to an intraabdominal complication such as anastomotic leakage, abscess, or local inflammation [15, 21]. Recent studies have shown that DGE does not occur as a result of pylorus preservation but rather as a consequence of postoperative complications [17, 22, 23]. Although the exact mechanism underlying DGE is not clear, our results suggest that DGE is related to clinical or even subclinical local inflammation caused by postoperative complications; three of our five patients with DGE (DGE14) had abscess or pancreatic fistula.

DGE has been generally defined as DGE10 (need for a nasogastric tube for more than 10 days) and DGE14 (failure to tolerate solid food by POD 14). The reported incidence of DGE ranges from 20 to 60% [5–13]. In the present study, no patient needed nasogastric decompression for more than 3 days. The nasogastric tube was removed on POD 1 in 30 (86%) of the 35 patients and on POD 2 in the remaining five. None required reinsertion of a nasogastric tube. With respect to DGE14, failure to tolerate solid food was observed in 5 of 35 patients, for an overall incidence of 14%.

A difference in DGE with respect to the reconstruction route, whether antecolic or retrocolic duodenojejunostomy, has been reported. In a retrospective study, Park et al. [23]. found that the incidence of DGE was 31.7% in the retrocolic group, but only 6.5% in the antecolic group. Hartel et al. [24]. reported an incidence of 24% with the retrocolic route and 5% with the antecolic route. Sugiyama et al. [25]. reported that DGE occurred in 1 of 12 patients (8%) in the antecolic group, but in 13 of 18 patients (72%) in the retrocolic group. These retrospective studies have suggested that the incidence of DGE is lower with the antecolic route than with the retrocolic route. A recent prospective randomized study by Tani et al. [18]. yielded an incidence of 50% for the retrocolic route, but 5% for the antecolic route. In the current prospective randomized

controlled trial, the incidence of DGE was 22% with the vertical retrocolic route and 6% with the antecolic route, but the difference was not statistically significant. Although the purpose of this study was to show the superiority of the vertical retrocolic route, an interim analysis did not show any advantage of the vertical retrocolic route; hence, we decided to terminate the study.

In addition to clinically defined DGE, the ^{13}C-acetate gastric emptying test showed that gastric emptying on POD 30 did not differ between the antecolic route and the vertical retrocolic route. Moreover, the gastric emptying did not recover to the preoperative level by 30 days in approximately 20% of patients, regardless of the reconstruction route. A greater increase in Tmax and T1/2 was observed with the vertical retrocolic route than with the antecolic route. These results suggest that the vertical retrocolic route offers no advantage. An analysis of the individuals showed that gastric emptying variables (Tmax, T1/2) had recovered to the preoperative level in approximately 80% of patients on POD 30, regardless of the reconstruction route. The day of analysis and type of meal selected (POD 30, liquid meal) should be reconsidered in another study.

In conclusion, the incidence of DGE and gastric emptying variables (Tmax, T1/2) after PpPD were similar between patients in whom reconstruction was performed by the antecolic route and those in whom it was performed by the vertical retrocolic route. On POD 30, gastric emptying was impaired in both groups compared to the preoperative level, but an analysis of individuals showed that it had recovered to the preoperative level in most patients, regardless of the reconstruction route.

Acknowledgments A part of this study was supported by grant-in-aid (No. 17591417) from Japanese Ministry of Education, Culture, Sports, Science and Technology. Dr. Imamura is an equally contributed first author.

References

1. Traverso LW, Longmire WP Jr. Preservation of the pylorus in pancreaticoduodenectomy. Surg Gynecol Obstet. 1978;146:959–62.
2. Kozuschek W, Reith HB, Waleczek H, Haarmann W, Edelmann M, Sonntag D. A comparison of long term results of the standard Whipple procedure and the pylorus preserving pancreatoduodenectomy. J Am Coll Surg. 1994;178:443–53.
3. Zerbi A, Balzano G, Patuzzo R, Calori G, Braga M, Carlo VD. Comparison between pylorus-preserving and Whipple pancreatoduodenectomy. Br J Surg. 1995;82:975–9.
4. Mosca F, Giulianotti PC, Balestracci T, Candio GD, Pietrabissa A, Sbarana F. Long-term survival in pancreatic cancer: pylorus-preserving versus Whipple pancreatoduodenectomy. Surgery. 1997;122:553–66.
5. Warshaw AL, Torchiana DL. Delayed gastric emptying after pylorus-preserving pancreaticoduodenectomy. Surg Gynecol Obstet. 1985;160:1–4.
6. Jiminez RE, Fernandez-del Castillo C, Rattner DW, Chang YC, Warshaw AL. Outcome of pancreaticoduodenectomy with pylorus preservation or with antrectomy in the treatment of chronic pancreatitis. Ann Surg. 2000;231:293–300.
7. Itani KM, Coleman RE, Meyers WC, Akwari OE. Pylorus-preserving pancreaticoduodenectomy. A clinical and physiological appraisal. Ann Surg. 1986;204:655–64.
8. Yeo CJ, Cameron JL, Sohn TA, Lillemoe KD, Pitt HA, Talamini MA. Six hundred fifty consecutive pancreaticoduodenectomies in the 1990s: pathology, complications, and outcomes. Ann Surg. 1997;226:248–57.
9. Fabre JM, Burgel JS, Navarro F, Boccarat G, Lemoine C, Domergue J. Delayed gastric emptying after pancreaticoduodenectomy and pancreaticogastrostomy. Eur J Surg. 1999;165:560–65.
10. Patel AG, Toyama MT, Kusske AM. Pylorus-preserving resection for pancreatic cancer. Is it any better? Arch Surg. 1995;130:838–43.
11. Braasch JW, Deziel DJ, Rossi RL, Watkins E Jr, Winter PF. Pyloric and gastric preserving pancreatic resection: experience with 87 patients. Ann Surg. 1986;204:411–8.
12. Yeo CJ, Barry K, Sauter PK, Sostre S, Lillemoe KD, Pitt HA, et al. Erythromycin accelerates gastric emptying after pancreaticoduodenectomy. Ann Surg. 1993;218:229–37.
13. Yamaguchi K, Tanaka M, Chijiiwa K, Nagakawa T, Imamura M, Takada T. Early and late complications of pylorus-preserving pancreatoduodenectomy in Japan 1998. J Hepatobiliary-Pancreat Surg. 1999;6(3):303–11.
14. Lin P, Lin YJ. Prospective randomized comparison between pylorus-preserving and standard pancreaticoduodenectomy. Br J Surg. 1999;86:603–7.
15. van Berge Henegouwen MI, van Gulik TM, de Wit LT, Allema JH, Rauws EA, Obertop H, et al. Delayed gastric emptying after standard pancreaticoduodenectomy versus pylorus-preserving pancreaticoduodenectomy: an analysis of 200 consecutive patients. J Am Coll Surg. 1997;185:373–9.
16. Neoptolemos JP, Russell RC, Bramhall S, Theis B. Low mortality following resection for pancreatic and periampullary tumours in 1026 patients: UK surgery of specialist pancreatic units. Br J Surg. 1997;84:1370–6.
17. Horstmann O, Markus PM, Ghadimi MB, Becker H. Pylorus preservation has no impact on delayed gastric emptying after pancreatic head resection. Pancreas. 2004;28:69–74.
18. Tani M, Terasawa H, Kawai M, Ina S, Hirono S, Uchiyama K, et al. Improvement of delayed gastric emptying in pylorus-preserving pancreaticoduodenectomy Results of a prospective, randomized, controlled trial. Ann Surg. 2006;243:316–20.
19. Chijiiwa K, Ohuchida J, Hiyoshi M, Nagano M, Kai M, Kondo K. Vertical retrocolic duodenojejunostomy decreases delayed gastric emptying after pylorus-preserving pancreatoduodenectomy. Hepato Gastroenterol. 2007;54:1874–7.
20. Braden B, Adams S, Duan LP, Orth KH, Maul FD, Lembcke B, et al. The [13C] acetate breath test accurately reflects gastric emptying of liquids in both liquid and semisolid test meals. Gastroenterology. 1995;108:1048–55.
21. Hocking MP, Harrison WD, Sninsky CA. Gastric dysrhythmias following pylorus-preserving pancreaticoduodenectomy. Possible mechanism for early delayed gastric emptying. Dig Dis Sci. 1990;35:1226–30.
22. Riediger H, Makowiec F, Schareck WD, Hopt UT, Adam U. Delayed gastric emptying after pylorus-preserving pancreatoduodenectomy is strongly related to other postoperative complications. J Gastrointest Surg. 2003;7:758–65.
23. Park YC, Kim SW, Jang JY, Ahn YJ, Park YH. Factors influencing delayed gastric emptying after pylorus-preserving pancreatoduodenectomy. J Am Coll Surg. 2003;196:859–65.

24. Hartel M, Wente MN, Hinz U, Kleef J, Wagner M, Muller MW, et al. Effect of antecolic reconstruction on delayed gastric emptying after the pylorus-preserving Whipple procedure. Arch Surg. 2005;140:1094–9.

25. Sugiyama M, Abe N, Ueki H, Masaki T, Mori T, Atomi Y. A new reconstruction method for preventing delayed gastric emptying after pylorus-preserving pancreatoduodenectomy. Am J Surg. 2004;187:743–6.

PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. 2008; 7: 202–214
Published online 6 June 2007 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

MAIN PAPER

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

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

Pharmaceutical STATISTICS

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

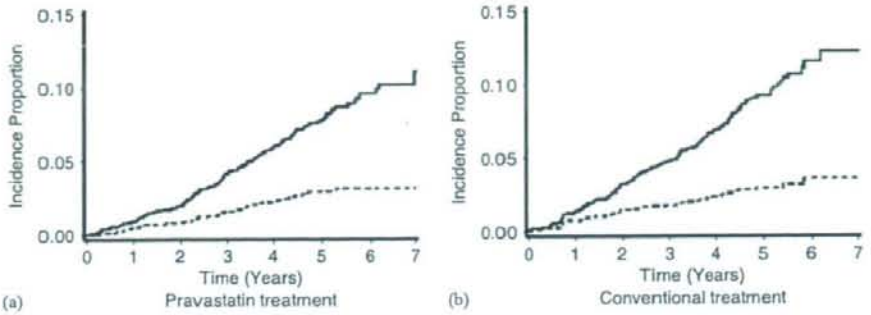

Figure 1. Peterson bounds for the marginal incidence proportion of CHD events in each treatment group. The solid line is one extreme scenario, in which subjects with competing events were assumed to experience CHD immediately after censoring. The dashed line is the other extreme scenario, in which subjects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

Pharmaceutical STATISTICS

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),

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

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 $\boldsymbol{\alpha}_{RC_j}$ 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,

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

where $\hat{\lambda}_{0RC_j}(X_u) = \sigma_{uj} / \sum_{i=1}^{n} \exp(\hat{\boldsymbol{\alpha}}_{RC_j} \overline{\mathbf{V}}_{iX_u}) Y_i(X_u) I(R_i = R)$ 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

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 $\overline{V}_t$. 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 $\overline{V}_t$ 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

Pharmaceutical STATISTICS

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 $25.9-$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

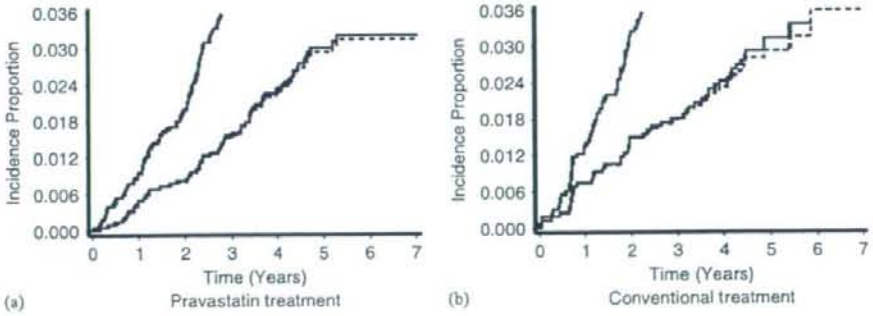

Figure 2. The IPCW marginal incidence proportion of CHD events in each treatment group. In each group, the upper solid line is the Peterson upper bound and the dashed line is the lower bound.

DOI: 10.1002/pst

Pharmaceutical STATISTICS

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.

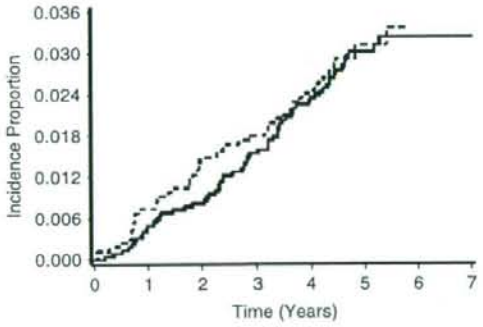

Figure 3. Comparison of the IPCW marginal incidence proportion of CHD events between treatment groups. Baseline confounding was adjusted by the IPTW method. The solid line is the pravastatin treatment group and the dashed line is the conventional treatment group.

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

aside from the competing risks problems, this final result depends heavily on the correct specification of the parametric model forms, which are usually unknown in most epidemiologic applications, while our IPTW adjustments of confounding do not need such parametric assumptions.

5. DISCUSSION

The problem of analyzing and interpreting data concerning competing risks continues to be one of the most important and vexing in biostatistical practice. The analyses of competing risks can be made using observable population parameters. An important observable quantity is the cumulative incidence functions based on the cause-specific hazards [25–28]. Alternatively, in this paper, we presented a method for estimation and comparison of treatment group-specific marginal survival curves of time to event data in the presence of dependent competing risks. The parameter of interest in our analysis is the marginal survival distribution, which is the net probability of time to event if only one cause of event acted on a population [4, 29, 30]. The ability to isolate the effect of one risk acting a population is attractive, especially if the focus of a study is to evaluate the effect of an intervention that is targeted at reducing incidence from that specific cause. Much of the literature on competing risks approaches such a problem by assuming the existence of latent survival times for each subject, that is, the estimation of event rates for certain types of event given the removal of some or all other event types. However, the net probabilities are hypothetical quantities and not directly observable in a population. Only observable quantities are their bounds, which allow for any possible dependence structure and will often be too wide to be of value [9–11]. Therefore, it is necessary to assume some model concerning the censoring process such as (1) to identify the net probabilities from the available information on the observables including covariate histories.

The proposed method is a straightforward extension of Robins and Finkelstein [14] for settings with two or more reasons for censoring. The application of the proposed methodology to the KLIS data suggested that the IPCW marginal incidence for CHD was almost the same as the lower bound. We included as many covariates as possible to predict the conditional probabilities of having remained uncensored. This result may suggest that there was little evidence of dependent competing risks in the KLIS data. In many studies, 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 are examined in detail and the effects on the final conclusion in the study should be discussed. On the other hand, when the results are nearly the same ones like the KLIS data, dependent censorings observed in the study does not cause a severe selection bias attributable to the covariates and the results from the standard analysis are robust in relation to the censorings.

It must be noted that the low incidences for the competing events do not always mean that the IPCW estimate will be close to the lower bound. The issue of interest in our analysis is whether the competing events are informative for their unobserved CHD events and whether the relation can be explained by the observed covariate histories. If they have much information on their unobserved CHD events and the covariates are available to explain the dependency, the IPCW estimate will be close to the upper bound without regard to the incidence of competing events. On the other hand, if they have little information on their unobserved CHD events, the IPCW estimate will be close to the lower bound.

For example, in a cohort of 100 subjects, suppose that at the time of 2 (years) from the start of follow-up, 5 subjects and 3 subjects experienced CHD events and non-CHD deaths, respectively, and that the remaining 92 subjects were censored at the end of study (time = 5). In this hypothetical data, the upper and lower bound

DOI: 10.1002/pst

Pharmaceutical STATISTICS

of the incidence rate is $8/(5 \times 2 + 3 \times 2 + 92 \times 5) = 1/59.5$ and $5/(5 \times 2 + 3 \times 5 + 92 \times 5) = 1/97$, respectively. Under the independent censoring, the incidence rate is $5/(5 \times 2 + 3 \times 2 + 92 \times 5) = 1/95.2$. If a binary covariate V is available and the distribution is $V = 1$ for both 5 CHD and 3 non-CHD events and $V = 0$ for 92 censored events at time = 5, the IPCW weights are 8/5 for the former events and 1 for the latter events. In this scenario of dependent competing risks, the IPCW incidence rate is $5 \times 1.6/(5 \times 2 \times 1.6 + 3 \times 2 \times 1.6 + 92 \times 5 \times 1) = 1/60.7$, which is almost the same as the upper bound. On the other hand, if the distribution of V is $V = 1$ for 5 CHD events and $V = 0$ for the other events, the IPCW weights are 1 for the former events, and are 95/92 (time $\leqslant 2$) and 1 ($2 <$ time $\leqslant 5$) for the latter events. In this scenario of independent competing risks, the IPCW incidence rate is $5 \times 1/(5 \times 2 \times 1 + 95 \times 2 \times 95/92 + 92 \times 3 \times 1) = 1/96.4$, which is almost the same as the lower bound. For more formal explanations, see Scharfstein and Robins [15] and Scharfstein *et al.* [31], in which the relation between the IPCW estimator and the bounds is discussed. They showed that, as the censoring bias parameters α in (2) goes to $\pm\infty$ (although they consider the case where the cause-specific hazard of censoring depends also on the possibly unobserved event time T given $\overline{V}_t$), the resulting IPCW estimator will converge to the bounds.

Our results are based on a nonidentifiable assumption concerning the residual dependence between time to events and competing risks due to unmeasured factors. The ordinary Kaplan–Meier estimator does not utilize recorded information on time-dependent covariates $\overline{V}_t$ and assumes the independence among competing risks, while the IPCW one utilizes such information and assumes the conditional independence among them. However, because causal interpretation of IPCW estimates depends on the correctness of assumption (1), making the censoring process ignorable is more important than fitting a parsimonious model in (2). As Joffe *et al.* [32] have described in the modeling of competing causes of death, the aggregation of censoring by competing causes may obscure important differences in the effect of various predictors on each type of censoring and so lead to misspecification of the model for censoring. Therefore, we fitted separate models for each type of censoring, where the treatment group-specific baseline hazard and regression parameters were assumed for each competing risk. Furthermore, in the KLIS, many clinically important time-dependent factors were measured and all of them were used as covariates to predict the probability of remaining in the study. Therefore, there will be a certain degree of validity in our IPCW estimates.

Otherwise, it will be necessary to develop sensitivity analysis methodology to investigate the sensitivity of our inferences to the fundamental assumption (1) of no unmeasured confounders. This sensitivity analysis will be particularly important for our data, in which the IPCW marginal incidence was almost the same as the lower bound. A simple and easy sensitivity analysis is to generate a hypothetical prognostic factor both for CHD and for competing events and to include the factor in the prediction of the conditional probabilities of uncensored. In the KLIS data, a hypothetical binary time-dependent covariate with a hazard ratio of 40.0 for both CHD and competing events was randomly generated and was included in the estimation of the subject-specific weight in addition to the 17 covariates described in Section 4. The increase of the resulting IPCW incidence in each group was slight compared with the estimates (Figure 2) ignoring the effect of the hypothetical unmeasured covariates. Therefore, in the KLIS data, it is likely that the effect of unmeasured confounders on our inferences would be small. Scharfstein *et al.* [15,31] have developed more formal sensitivity analysis. The sensitivity analysis for our data using their idea will be future work.

ACKNOWLEDGEMENTS

This research was supported in part by a Grant-in-Aid for Scientific Research (A) No. 16200022. We would like to thank the Kyushu Lipid Intervention Study Group (Principal Investigator: Dr Kikuo Arakawa) for permission to use these valuable data. We also thank two anonymous reviewers for their comments, which led to a much improved version of the paper.

DOI: 10.1002/pst

REFERENCES

1. The Kyushu Lipid Intervention Study Group. A coronary primary intervention study of Japanese men: study design, implementation and baseline data. *Journal of Atherosclerosis and Thrombosis* 1996; 3:95–104.
2. The Kyushu Lipid Intervention Study Group. Pravastatin use and risk of coronary events and cerebral infarction in Japanese men with moderate hypercholesterolemia: the Kyushu Lipid Intervention Study. *Journal of Atherosclerosis and Thrombosis* 2000; 7:110–121.
3. Gail M. A review and critique of some models used in competing risk analysis. *Biometrics* 1975; **31**: 209–222.
4. Kalbfleish JD, Prentice RL. *The Statistical Analysis of Failure Time Data* (2nd edn). Wiley: New York, 2002.
5. Tsiatis AA. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* 1975; **72**:20–22.
6. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; **53**:457–481.
7. Benfante R, Yano K, Hwang LJ, Curb D, Kagan A, Ross W. Elevated serum cholesterol is a risk factor for both coronary heart disease and thromboembolic stroke in Hawaiian Japanese men: implications of shared risk. *Stroke* 1994; **25**:814–820.
8. Peterson AV. Bounds for a joint distribution function with fixed subdistribution functions: applications to competing risks. *Proceedings of the National Academy of Sciences* 1976; **73**:11–13.
9. Slud EV, Rubinstein LV. Dependent competing risks and summary survival curves. *Biometrika* 1983; **70**:643–649.
10. Klein JP, Moeschberger ML. Bounds on net survival probabilities for dependent competing risks. *Biometrics* 1988; **44**:529–538.
11. Zheng M, Klein JP. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* 1995; **82**:127–138.
12. 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.
13. 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.
14. 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.
15. Scharfstein DO, Robins JM. Estimation of the failure time distribution in the presence of informative censoring. *Biometrika* 2002; **89**:617–634.
16. Robins JM. Causal inference from complex longitudinal data. *Latent variable modeling and application to causality*, Berkane M (ed.). Springer: New York, 1997; 69–117.
17. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**:41–55.
18. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 1952; **47**:663–685.
19. Satten GA, Datta S. The Kaplan–Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician* 2001; **55**: 207–210.
20. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*. University of California Press: Berkeley, 1976; 221–233.
21. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* 1989; **84**:1074–1078.
22. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**:550–560.
23. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, Halloran ME Berry DA (eds). Springer: New York, 1999; 95–133.
24. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology* 2003; **14**:680–686.
25. Gray RJ. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics* 1988; **16**:1141–1154.
26. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**: 496–509.
27. Anderson PK, Abildstrom SZ, Rosthoj S. Competing risks as a multi-state model. *Statistical Methods in Medical Research* 2002; **11**:203–215.
28. Anderson PK, Keiding N. Multi-state models for event history analysis. *Statistical Methods in Medical Research* 2002; **11**:91–115.
29. Tsiatis AA. Competing risks. *Encyclopedia of Biostatistics*, vol. 1, Armitage M, Colton T (eds). Wiley: New York, 1997; 824–834.

DOI: 10.1002/pst

Pharmaceutical STATISTICS

30. Crowder MJ. *Classical Competing Risks*. Chapman & Hall/CRC: London, 2001.
31. Scharfstein DO, Robins JM, Eddings W, Rotnitxky A. Inference in randomized studies with informative censoring and discrete time-to-event endpoints. *Biometrics* 2001; **57**:404–413.
32. Joffe MM, Hoover DR, Jacobson LP, Kingsley L, Chmiel JS Visscher BR, Robins JM. Estimating the effect of zidovudine on kaposi's sarcoma from observational data using a rank preserving structural failure-time model. *Statistics in Medicine* 1998; **17**:1073–1102.

DOI: 10.1002/pst

STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:1718–1733
Published online 5 October 2007 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