Complete pancreatic resection has been able to yield actuarial 5-year survival rates of 15%–25% following pancreaticoduodenoctomy[2–4] and 8%–14% following distal pancreatectomy.[5,6] The common failure patterns in treating pancreatic cancer have been recurrence in the locoregional tumor bed and/or the development of hepatic metastases. We introduced extended radical pancreatectomy combined with intraoperative radiation therapy (IORT) in 1984. Compared with other treatment modalities, this approach has provided the best control of local recurrence,[7] although it has not contributed to survival benefit. The cumulative survival curve after surgery for advanced pancreatic cancer is characterized by a steep decline in the early postoperative period followed by a gentle downward slope. This indicates that most patients die not long after resection. Our aim is to identify the characteristics associated with early mortality in patients undergoing pancreatic resection for pancreatic cancer.

PATIENTS AND METHODS

Patients and Surgical Technique

Between December 1984 and December 1999 at Kumamoto University, we performed pancreatic surgery on 88 patients with pancreatic cancer. Of the 88 patients, 41 underwent extended radical pancreatectomy combined with IORT, because they had potentially curable disease as assessed by physical examination and the following objective imaging criteria: (1) no evidence of remote metastases; (2) no evidence of tumor extension to the celiac axis or the superior mesenteric artery; (3) possible resection and reconstruction of the superior mesenteric vein (SMV) or the portal vein (PV) in cases of tumor extension to the SMV or PV. This combined therapy, which had provided the best control of local recurrence, involves the dissection of the juxtaparaaortic and regional lymph nodes, together with the connective tissue and nervous plexus around the aorta, extending from the diaphragm above to the inferior mesenteric artery below. After dissection, a dose of 30 Gy with 9–12 MeV electron beam radiation was administered to the operative field, as described previously.[7] Four cases of hospital death were excluded from this study. Twelve patients received preoperative and/or postoperative 5-fluorouracil (5-FU)-based chemotherapy. No patients received extra radiation therapy.

Survival Curve Analyzed

The cumulative survival curve in this series was depicted using the Kaplan-Meier method. Assuming that there were two distinct arms of the survival curve below and above the breakpoint, each part of the curve was modeled as an exponential distribution. The exponential model corresponded to an assumption of a uniform hazard within that time period. Three parameters, the breakpoint, the high hazard rate below the breakpoint, and the low hazard rate above the breakpoint were estimated by the maximum likelihood method. The model fitness, the magnitude of −2 log likelihood was also evaluated, changing the breakpoint and the case having no breakpoint. The better fit model had a lower value of −2 log likelihood. The short survival group (SSG) was defined based on analyses of the survival curve using these statistical methods.

Characteristics Analyzed

Prognostic factors associated with early mortality after surgery were evaluated for the following variables: clinical and histological features, laboratory examination data, and intraoperative factors. The analyzed factors were as follows: clinical features (abdominal pain and/or back pain, body weight loss, preoperative obstructive jaundice and/or biliary infection); histological features (tumor location, tumor size, tumor stage, residual tumor); laboratory examination data (white blood cell, lymphocytes, hemoglobin, total protein, albumin, serum amylase level, elastase-1, the 75 g glucose tolerance test, the N-benzoyl-L-tyrosyl-P-aminobenzoic acid (NBT-PABA) test, tumor-associated carbohydrate antigens, CA 19-9 and/or DUPAN-2 and/or Span-1); intraoperative factors (operation time, bleeding volume, volume of blood transfusion). We conducted an interview with each patient during clinical course to inquire about the experience of abdominal pain and/or back pain and body weight loss. Preoperative biliary drainage (endoscopic stents or percutaneous drains) was performed in the event of biliary infection and/or jaundice (serum total bilirubin level more than 5 mg/dl). Blood samples for the complete blood count, chemistries, amylase, and tumor-associated carbohydrate antigens were taken on admission. All but one patient were measured for serum tumor-associated carbohydrate antigens. Within the normal limits of serum CA 19-9, DUPAN-1, and Span-1 levels are 37 U/ml, 150 U/ml, and 30 U/ml, respectively. The 75 g glucose tolerance test and the NBT-PABA test were performed before operation. Diagnosis of glucose intolerance was made on the basis of a

fasting venous blood glucose level of more than 110 mg/dl, and/or glucose level more than 140 mg/dl 2 hours after uptake of 75 g glucose. The exocrine function of the pancreas was assessed by the NBT-PABA test. Insufficiency of exocrine function was diagnosed by lower than 70% excretion of total PABA in the urine. Histopathological examination of each resected specimen was conducted. The resected specimens were fixed in a 20% formalin solution. The paraffin sections for histological examination were prepared and stained with hematoxylin and eosin. According to the classification of pancreatic cancer defined by the Japanese Pancreas Society,[8] we evaluated histological characteristics including tumor location, tumor size, tumor stage, and residual tumor.

Statistical Analysis of Prognostic Factors

Prognostic factors associated with early mortality after surgery were first investigated using univariate Cox proportional hazards regression analyses, and significant independent variables were subsequently tested in a multivariate Cox proportional hazards regression analysis. Statistical difference was considered to be significant at $P < 0.05$.

RESULTS

Patient Characteristics

The clinical characteristics of the 37 patients are outlined in Table 1. Study subjects included 15 women and 22 men. The average age of the patients was 59.1 years, ranging from 37 to 75 years. The primary pancreatic lesion was located in the head in 30 patients and in the body in 7 patients. Of these 37 patients, 28 were affected by stage IVa disease. Twenty-three patients underwent R0 resection.

Analyses of Survival Curve

The cumulative survival curve in this series was depicted using the Kaplan-Meier method. Five-year actual survival was 14.0%. This Kaplan-Meier survival curve could not be approximately represented by an exponential distribution with the constant hazard rate of 0.028 (−2 log likelihood was 311.851) if there was no breakpoint. Assuming that there are two distinct arms of the survival curve, below and above the breakpoint, each part of the curve was modeled as the exponential distribution. Three parameters, the breakpoint, the high hazard rate below the breakpoint, and the low hazard rate above the breakpoint were estimated by the maximum likelihood method. The magnitude of −2 log likelihood was also evaluated, changing the breakpoint and the cases having no breakpoint. Using the maximum likelihood estimate, if the breakpoint was 41, the value of the −2 log likelihood was minimal (300.072) (Table 2). Therefore, we estimated the breakpoint at 41 months. This Kaplan-Meier survival curve could be approximately represented by two distinct exponential curves with the different hazard rates below and above the breakpoint (Fig. 1). The hazard rate of the survival curves below and above the breakpoint was 0.038 and 0.0072, respectively. The difference of −2 log likelihood between the constant hazard rate model and the two distinct hazard rate model was 11.779, and it was highly significant ($P = 0.003$).

According to these analyses, SSG was defined as death earlier than 41 months after surgery, including 31 patients (73.0%). The long surviving patient group (LSG) consisted of 6 patients who were still alive more than 41 months postoperatively.

Table 1.
Patients characteristics

No. of patients	37
Age (years)	
Average	59.1
Range	37–75
Male/female	22/15
Site of primary lesion	
Head	30
Body	7
Tail	0
Tumor stage	
I	0
II	1
III	5
IVa	28
IVb	3
Pancreatectomy	
PpPD	11
PD	17
DP	6
TP	3

PpPD, pylorus preserving pancreaticoduodenectomy; PD, pancreaticoduodenectomy; DP, distal pancreatectomy; TP, total pancreatectomy.

Outcomes

All 37 patients were followed until death, or from 56 to 234 months for surviving patients. The median follow-up for surviving patients is 82.7 months. The outcomes of

Table 2.
Maximum likelihood estimate

Break point (BP)	20	30	41*	50	60	none
−2 log likelihood	311.381	307.246	300.072	302.753	305.391	311.351
Hazard rate Before BP	0.0312	0.0359	0.0382	0.0381	0.0341	0.0278
Hazard rate After BP	0.0246	0.0159	0.0072	0.0081	0.0094	0.0278

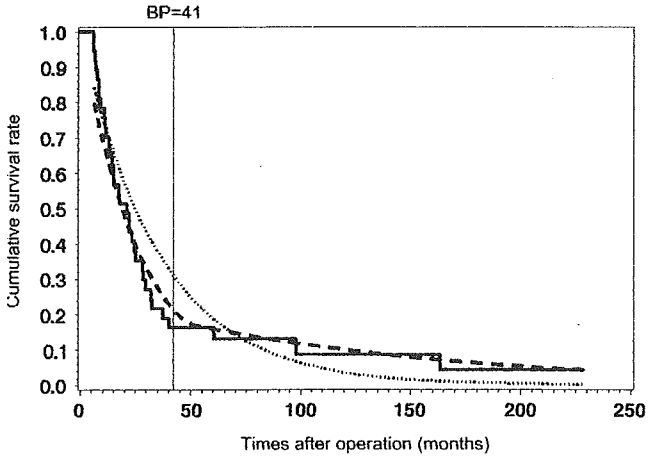

Figure 1. The cumulative survival curve in this series was depicted using the Kaplan-Meier method (a solid line). This curve could not be approximately represented by an exponential curve with a hazard rate of 0.028, if there was no breakpoint (a light dotted line), but it could be approximately represented by two distinct arms of the exponential curves that composed the survival curves below and above the breakpoint (bold dotted line). BP: breakpoint.

these patients are described in Table 3. In SSG, there were 29 (95.8%) cases of cancer-related death, compared with 5 such cases (33.3%) in LSG. The most striking outcome was that 18 patients (58.1%) died of hepatic metastases in SSG, whereas no patients died of hepatic metastases in LSG.

Prognostic Factors Associated with Early Mortality

First, using a Cox proportional hazards model, a univariate analysis was performed to determine whether any of the characteristics were predictors of survival. Among clinical and pathological variables, abdominal pain and/or back pain during the preoperative clinical course was the only significantly prognostic factor (Table 4). Among laboratory data, preoperative elevation of serum amylase, abnormality in NBA-PABA test, and elevation of tumor-associated carbohydrate antigens levels were statistically significant prognostic factors (Table 5). The intraoperative factors, including operation time, bleeding volume, and blood transfusion were also analyzed. No significant predictive factor for early death was identified among the intraoperative variables (Table 6).

Table 3.
Mode of recurrence in patients with extended pancreatic resection with IORT

	SSG (31 patients)	LSG (6 patients)
Cancer-related Death		
Liver metastases	18	0
Lung metastases	2	1
Peritoneal dissemination	6	0
Pleural dissemination	0	1
Recurrence of remnant pancreas	2	0
Local recurrence	1	0
Other cause of death	2	1
Alive	0	3

SSG, short survival group; LSG, long surviving patients group; IORT, intraoperative radiation therapy.

Table 4.
Univariate analysis of prognostic factors associated with early mortality; clinical and pathological characteristics

	Hazard ratio	95% CI	*P* value
Abdominal pain and/or back pain	3.216	1.517–6.817	0.0023
Body weight loss	1.330	0.660–2.681	0.4252
Obstructive jaundice and/or cholangitis	1.534	0.748–3.147	0.2434
Tumor location (head)	1.118	0.481–2.596	0.7957
Tumor size (cm)	1.073	0.886–1.299	0.4691
Tumor stage IV	2.107	0.735–6.045	0.1655
Residual tumor (R1-2)	0.803	0.395–1.632	0.5438

A Multivariate Analysis of Prognostic Factors

A multivariate Cox proportional hazards regression analysis of the factors that significantly predicted survival in the univariate analysis revealed that abdominal pain and/or back pain during the preoperative clinical course was the only prognostic factor (Table 7).

Table 5.
Univariate analysis of prognostic factors associated with early mortality; laboratory data

	Hazard ratio	95% CI	*P* value
White blood cell	1.000	1.000–1.000	0.7187
Lymphocytes	1.000	1.000–1.001	0.4768
Hemoglobin	1.293	0.984–1.699	0.0649
Total protein	1.051	0.563–1.961	0.8766
Albumin	1.339	0.566–3.169	0.5067
Elevation of serum Amylase level	3.213	1.488–6.937	0.0030
Elevation of Elastase-1	1.674	0.718–3.904	0.2332
Abnormal glucose metabolism	0.940	0.459–1.925	0.8652
NBT-PABA test	2.613	1.165–5.861	0.0198
tumor-associated carbohydrate antigens	2.974	1.179–7.501	0.0209

Table 6.
Univariate analysis of prognostic factors associated with early mortality; intraoperative factors

	Hazard ratio	95% CI	*P* value
Operation time (hr)	1.089	0.935–1.269	0.2743
Bleeding volume (L)	1.204	0.886–1.636	0.2343
Blood transfusion (unit)	1.025	0.998–1.052	0.0671

DISCUSSION

Pancreatic adenocarcinoma remains a lethal disease. A small coterie of patients can be salvaged, at least temporarily, by surgical resection. Compared with other treatment modalities, extended radical pancreatectomy combined with IORT has provided the best control of local recurrence,[7] although it does not appear to have made a contribution to survival benefit. The most crucial problem, apparent in more than half the patients in this series, is occurrence of hepatic metastases after operation, despite the absence of any evidence of liver metastases by preoperative radiological imaging and intraoperative examination. Even after curative resection of 21 patients with pancreatic cancer, liver metastasis occurred in 11 patients (52.4 %).[9] Liver metastasis occurred in a total of 18 cases (48.6 %) in this study, and these patients died of hepatic metastases within 41 months after operation. It is suggested that the steep downward slope of the survival curve in the early postoperative phase reflects the poor prognosis of patients who died of hepatic metastases.

Generally, survival curves of malignant diseases with poor prognosis do reveal a steep downward slope in the early postoperative phase. This implies that improvement of treatment causes the steep downward slope to change to a gentle downward slope in the survival curve. Improvement of short-term survival for pancreatic cancer can be made by establishing means of preventing liver metastases, but two crucial factors must be considered. First, is the need for more precise tools for preoperative diagnosis of liver metastases, because the high recurrence rate of liver metastases in the early period after surgery indicates that liver micro-metastases might be present at the time of operation. One advantageous diagnostic tool to detect small liver metastases is computed tomography during arterial portography combined with computed tomography–assisted hepatic arteriography (CTAP + CTHA). CTAP + CTHA has a higher sensitivity than conventional CT for detection of liver metastases.[10] The second factor is treatment for free cancer cells in the blood and latent liver micro-metastases that may not be detected by preoperative radiological imaging and intraoperative examination. Analyses of slope of survival curve in this study can be useful to evaluate the efficacy of current and future treatment for liver metastases after surgery.

Table 7.
Multivariate analysis of prognostic factors associated with early mortality

	Hazard ratio	95% CI	*P* value
Abdominal pain and/or back pain	3.780	1.431–9.988	0.0073
Elevation of serum Amylase level	1.630	0.619–4.289	0.3226
NBT-PABA test	1.781	0.662–4.793	0.2528
Tumor-associated carbohydrate antigens	1.870	0.597–5.858	0.2825

Prognostic factors associated with early mortality after surgery were evaluated using univariate and multivariate Cox proportional hazards regression analyses. If three patients who had other causes of death were considered censored, there was no difference in the statistical results. From the multivariate Cox proportional hazards regression analysis, abdominal and/or back pain was found to be the only statistically significant prognostic factor in this study. The causes of abdominal and/or back pain might be the occurrence of pancreatitis or invasion of tumor into the retroperitoneal nerve. From the univariate analysis, the preoperative serum amylase level was found to be one of the prognostic factors. These findings suggest that inflammation caused by obstruction of the main pancreatic duct and its branches may promote rapid tumor progression. The causative link between

inflammation and cancer has been described, with evidence presented for several cancers of the gastrointestinal tract.[11] We have reported that a human pancreatic cancer cell line, Capan-1, expresses the chemokine receptor 2, which is an interleukin 8 (IL-8) receptor.[12] Interleukin-8 might contribute to tumor progression via nuclear factor kappa B (NFκB) activation, because IL-8 activates NFκB. Overexpression of downstream genes of NFκB, such as urokinase plasminogen activator, are important in cancer metastasis.[13] Moreover, pancreatitis also promotes elevation of serine protease level in the pancreas and blood. It was reported that serine protease enhanced blood-borne metastases of tumor cells due to tumor cell aggregation.[14] This evidence suggests that there might be a strong correlation between inflammation and tumor progression of pancreatic cancer.

Elevation of tumor-associated carbohydrate antigen levels was also a predictive factor after surgery in this study under a univariate analysis. Sialyl Lewis^a antigen (CA 19-9) was reported to play an important role in the endothelial leukocyte adhesion molecule-1 mediated binding between human cancer cells and activated endothelial cells.[15] It has been suggested that tumor cells expressing Sialyl Lewis^a antigen have a high degree of adhesion to endothelial cells in the process of metastases. There are also reports of a positive correlation between expression of the tumor-associated carbohydrate antigens and the hepatic metastatic potential of pancreatic cancer.[9,16] These carbohydrate antigens might facilitate metastasis of cancer cells to the liver.

In conclusion, the Kaplan-Meier survival curve after surgery for advanced pancreatic cancer could be approximately represented by two distinct exponential curves, with the different hazards rates below and above the breakpoint. The high hazard rate in the early postoperative period was closely linked with death from liver metastases. Therefore, prevention and treatment of liver metastases should be a goal in improving survival of pancreatic cancer patients. The effects of prevention and treatment can be evaluated by analysis of the survival curve in this study. The preoperative presence of local pain, which may be caused by inflammation, was a prognostic factor associated with early mortality. Thus the relation between this factor and tumor progression should be precisely analyzed.

REFERENCES

1. Kern S, Hruban R, Hollingsworth MA, *et al.* A white paper: the production of a pancreas cancer think tank. Cancer Res 2001;61:4923–4932.
2. Baumel H, Huguier M, Manderscheid JC, *et al.* Results of resection for cancer of the exocrine pancreas: a study from the French association of surgery. Br J Surg 1994;81: 102–107.
3. Sohn TA, Yeo CJ, Cameron JL, *et al.* Resected adenocarcinoma of the pancreas—616 patients: results, outcomes, and prognostic indications. J Gastrointest Surg 2000;4:567–579.
4. Richter A, Niedergethmann M, Sturm JW, *et al.* Long-term results of partial pancreaticoduodenoctomy for ductal adenocarcinoma of the pancreatic head: 25-year experience. World J Surg 2003;27:324–329.
5. Dalton RR, Sarr MG, van Heerden JA, *et al.* Carcinoma of the body and tail of the pancreas: is curative resection justified? Surgery 1992;111:489–494.
6. Brennan MF, Moccia RD, Klimstra D. Management of adenocarcinoma of the body and tail of the pancreas. Ann Surg 1996;223:506–512.
7. Hiraoka T, Uchino R, Kanemitsu K, *et al.* Combination of intraoperative radiation with resection of cancer of the pancreas. Int J Pancreatol 1990;7:201–207.
8. Japan Pancreas Society (1996) Classification of Pancreatic Cancer. First English Edition, Kanehara & Co., Ltd, Tokyo, pp 40–43.
9. Takamori H, Hiraoka T, Kanemitsu K, *et al.* Pancreatic liver metastases after curative resection combined with intraoperative radiation for pancreatic cancer. Hepato-Gastroenterology 2004;51:1500–1503.
10. Takamori H, Ikeda O, Kanemitsu K, *et al.* Preoperative detection of liver metastases secondary to pancreatic cancer. Utility of combined helical computed tomography during arterial portography with biphasic computed tomography–assisted hepatic arteriography. Pancreas 2004;29:188–192.
11. Farrow B, Sugiyama Y, Chen A, *et al.* Inflammatory mechanisms contributing to pancreatic cancer development. Ann Surg 2004;239:763–771.
12. Takamori H, Oades ZG, Hoch RC, *et al.* Autocrine growth effect of IL-8 and GRO α on a human pancreatic cancer cell line, Capan-1. Pancreas 2000;21:52–56.
13. Wang W, Abbruzzese JL, Evans DB, *et al.* Overexpression of urokinase-type plasminogen activator in pancreatic adenocarcinoma is regulated by constitutively activated RelA. Oncogene 1999;18:4554–4563.
14. Sugihara S, Yamamoto T, Tsuruta J, *et al.* Serine protease-induced enhancement of blood-borne metastases of rat ascites tumour cells and its prevention with deoxyribonuclease. Br J Cancer 1990;62:607–613.
15. Takada A, Ohmori K, Takahashi N, *et al.* Adhesion of human cancer cells to vascular endothelium mediated by carbohydrate antigens, sialyl Lewis^a. Biochem Biophys Res Commun 1991;179:713–716.
16. Takamori H, Hiraoka T, Yamamoto T. Expression of tumor-associated carbohydrate antigens correlates with hepatic metastasis of pancreatic cancer: clinical and experimental studies. Hepato-Gastroenterology 1996;43:748–755.

A Simulation Study for a Linear Measurement Error Model When Error Variances Vary Between Measurements

Yasutaka CHIBA, Yutaka MATSUYAMA, Tosiya SATO, and Isao YOSHIMURA

When predicting scores in the Draize eye irritation test based on measurements of in vitro alternative tests, we are often faced with estimating parameters in a linear measurement error model with heterogeneous error variances. This article proposes a new statistical method for parameter estimation to address this issue. The proposed method is an extension of an earlier proposal that applied a linear measurement error model with homogeneous error variances, to cases with heterogeneous error variances. A simulation study to examine the performance of the proposed method was conducted in a framework that was adaptable to the data, which was obtained in a validation study of alternative methods to animal experiments conducted in Japan. The proposed method reduced the biases of estimates in comparison with an ordinary regression analysis method and three other methods under the assumption of homogeneous error variances. Although the proposed method did not fit the real data well, the resulting prediction formula was far better than those obtained by other methods.

Key Words: Alternative method to animal experiments; Chorioallantoic membrane assay; Draize eye irritation test; Estimating function; Functional relationship model.

1. INTRODUCTION

In the development of cosmetics, evaluation of eye irritancy caused by ingredients or metabolites is indispensable. Although the Draize rabbit eye irritation test (Draize test) proposed by Draize, Woodard, and Galvery (1944) has been conventionally used for this purpose, its use has been restricted recently due to mounting concerns about cruelty to animals. In the recent years, many methods that do not use animals have been developed (see, e.g., OECD 1996). One such method is a chorioallantoic membrane assay by trypan blue staining test (CAM-TB).

Yasutaka Chiba is a Doctorate Student, Department of Biostatistics, Kyoto University School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan (E-mail: chibay@pbh.med.kyoto-u.ac.jp). Yutaka Matsuyama is Associate Professor, Department of Biostatistics, Kyoto University School of Public Health. Tosiya Sato is Professor, Department of Biostatistics, Kyoto University School of Public Health. Isao Yoshimura is Professor, Department of Management Science, Tokyo University of Science, Kagurazaka 1-3, Shinjuku-ku, Tokyo 162-8601, Japan.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 10, Number 1, Pages 118–130
DOI: 10.1198/108571105X28679

An interlaboratory study to validate alternative methods including CAM-TB to the Draize test was conducted. In the course of establishing a formula to predict Draize test scores based on an alternative method such as the CAM-TB, a statistical issue arose. One approach to establishing a prediction formula is to demonstrate a functional relationship between the scores in the Draize test and the measurements in an alternative method. Each of the paired measurements (the CAM-TB and the Draize test) was measured with error. These measurements were governed by heterogeneous errors. This kind of problem often arises in regression analysis with measurement error models (e.g., Fuller 1987) when error variances are heterogeneous between measurements. This article proposes a new method for analyzing these types of data.

In Section 2, the data obtained from the Draize test and the CAM-TB is described. In Section 3, we formulate the problem for analyzing these data and propose a statistical method to solve this problem. In Section 4, we conduct a simulation study to examine the performance of the proposed method in comparison with other methods, which do not assume heterogeneous error variances. The simulation study was conducted in a framework that was adaptable to the data obtained in our application. In Section 5, we analyze the data described in Section 2. Finally, we discuss the implications of the proposed method.

2. FORMULATION OF PROBLEM

The Draize test is a test designed to predict the eye irritation potential of chemical substances (Draize et al. 1944). The focus of this test is an assessment of observable mucosal and epithelial effects. Test scores represent damage caused to a rabbit's eyes on a scale from 0 to 110. See Wilhelmus (2001) for a more detailed explanation.

Several parties, including scientists and the Society for the Prevention of Cruelty to Animals, actively criticize the use of animal experiments based on the belief that experiments such as the Draize test are inhumane. Accordingly, there is growing demanding for alternative methods to animal experiments.

The CAM-TB is a test devised as an alternative method to the Draize test (Hagino et al. 1991, 1993, 1999). A chorioallantoic membrane (CAM) assay evaluates blood vessel reaction and damage to the CAM of a fertilized hen's egg. A hen's egg is recognized as experiment material that lies between in vivo and in vitro. Consequently, the CAM assay is considered to be less cruel than the Draize test.

In order to identify effective alternative methods, the Japan Cosmetic Industry Association (1994) conducted a validation study. In this study, the Draize test and the CAM-TB were carried out using 36 chemical substances. For each chemical substance, three rabbits and five eggs were examined. The results provided three Draize test scores and five CAM-TB measurements for each chemical substance.

Figure 1 shows a scatterplot of the scores in the Draize test by the measurements in the CAM-TB, where dots denote means and whiskers denote standard deviations. Figure 2 shows the mean SD plots, which show the association between means and standard deviations for the Draize test and the CAM-TB. In the Draize test, error variances of scores

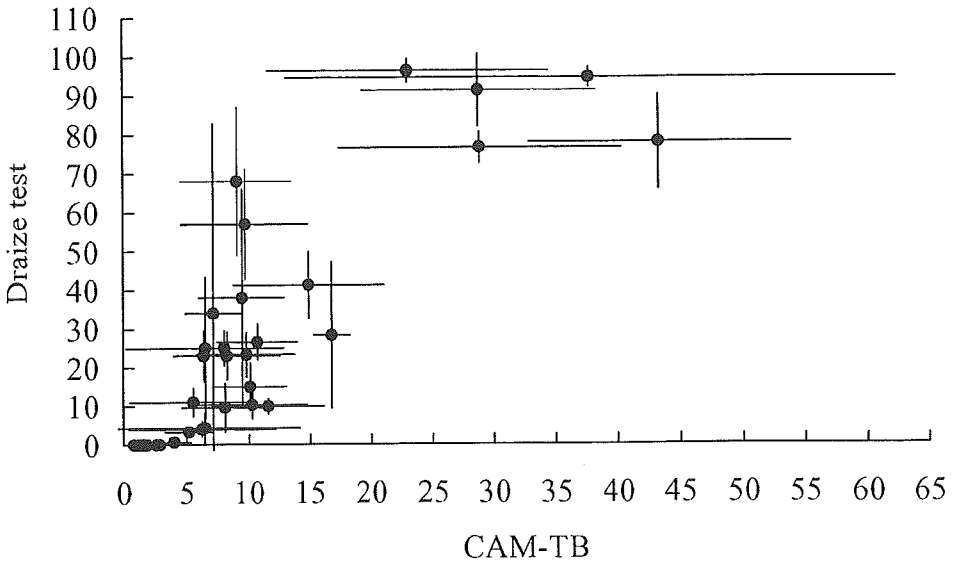

Figure 1. Scatterplot of the Draize Test and the CAM-TB: The vertical axis is the mean score of the Draize test and the horizontal axis is the mean of the CAM-TB, and whiskers express standard deviations.

for nonirritants or severe irritants were generally smaller than those for moderate irritants (top of Figure 2). In the CAM-TB, error variances of measurements increase in relation to the level of severity of irritants (bottom of Figure 2). Thus, we assume that error variance is dependent on the mean with a certain relationship. Because scores in the Draize test have an upper limit (110) and a lower limit (0), it is reasonable to fit a curve with sigmoid shape. Of all the curves with sigmoid shape, we choose the logistic curve because it is flexible and relatively easy to work with. We assume that the relationship between true scores in the Draize test, η, and true measurements in the CAM-TB, ξ, is well modeled by the logistic curve, $\eta = 110\exp(\beta_0 + \beta_1\xi) / \{1 + \exp(\beta_0 + \beta_1\xi)\}$. These actual data are measured with error. The principal problem of the validation study then lies in the evaluation of predictability of the CAM-TB to the Draize test. A suitable statistical method is required to estimate the regression parameters on a measurement error model with heterogeneous error variances. We employ the logit transformation, $\eta^* = \log\{\eta/(110 - \eta)\}$, which yields a linear relationship, to derive a suitable statistical method. This situation is formulated as follows.

Let $(x_i, z_i), i = 1, 2, \ldots, n$, denote a pair of measurements, and ξ_i, which is treated as a nuisance parameter, denote an unknown true value of x_i. A functional relationship exists between x_i's and z_i's such that

$$\begin{aligned} x_i &= \xi_i + \delta_i \\ z_i &= \beta_0 + \beta_1\xi_i + \varepsilon_i, \end{aligned}$$

where β_0 and β_1 are regression parameters, δ_i is the measurement error, and ε_i is the disturbance term in a linear regression model. It is assumed that δ_i and ε_i are distributed

according to

$$\begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{xi}^2 & 0 \\ 0 & \sigma_{zi}^2 \end{pmatrix} \right).$$

It is necessary to estimate these variances.

3. PARAMETER ESTIMATION METHOD

In order to find a less-biased parameter estimation method for the model derived in Section 2, we considered expanding Amari and Kawanabe's proposal (1997). They

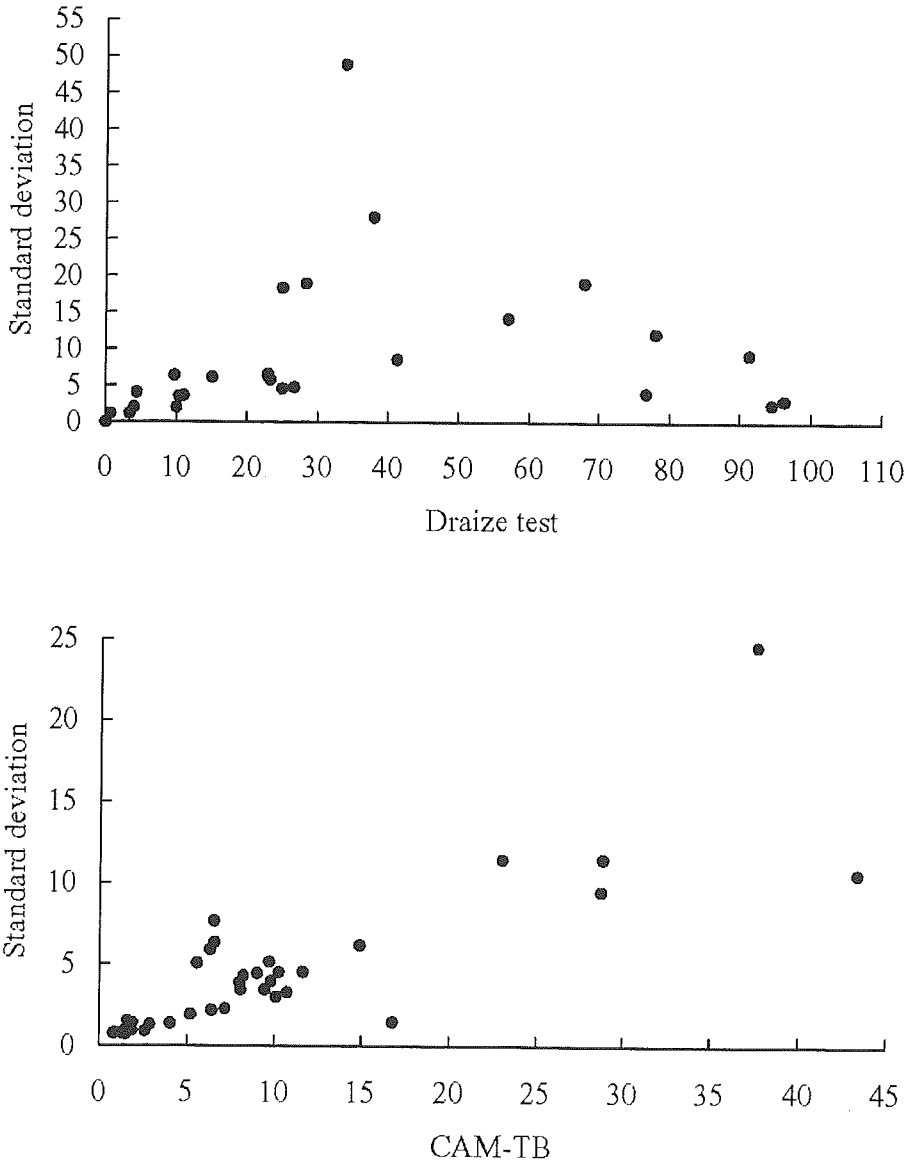

Figure 2. Top: Mean SD plot of the Draize test (vertical axis is standard deviation and horizontal axis is mean), Bottom: Mean SD plot of the CAM-TB (vertical axis is standard deviation and horizontal axis is mean).

discussed the problem that nuisance parameters increase with an increase in sample size. In some cases in which such problems arise, the maximum likelihood method does not derive consistent estimators (e.g., see Neyman and Scott 1948). The statistical analysis of linear measurement error models poses the same problem, that is, nuisance parameters increase with the increase of sample size. For this problem, Amari and Kawanabe (1997) defined a family of estimating functions under the assumption of homogeneous error variances. We extend their method to the case with heterogeneous error variances, and propose a parameter estimation method that makes asymptotic variances of parameters smallest in an extended family of estimating functions.

Under the common known variances, $\sigma_{xi}^2 = \sigma_{zi}^2 = \sigma^2$, Amari and Kawanabe (1997) showed that the estimating functions for β_0 and β_1, $g_0(x, z, \beta)$ and $g_1(x, z, \beta)$, become

$$g_0(x, z, \beta) = \frac{z - \beta_0 - \beta_1 x}{(1 + \beta_1^2)\sigma^2} \tag{3.1}$$

and

$$g_1(x, z, \beta) = h(s)\frac{z - \beta_0 - \beta_1 x}{(1 + \beta_1^2)\sigma^2}, \tag{3.2}$$

where $s = \{x + \beta_1(z - \beta_0)\}/(1 + \beta_1^2)$ is the sufficient statistic for ξ, and $h(s)$ is an arbitrary function of s. The estimating functions (3.1) and (3.2) can produce consistent estimators without depending on nuisance parameters $\xi_1, \ldots, \xi_n$. The asymptotic variance of $\beta_j, j = 0, 1$, a.v.$[\sqrt{n}(\hat{\beta}_j - \beta_j)]$, is given by

$$\text{a.v.}\left[\sqrt{n}\left(\hat{\beta}_j - \beta_j\right)\right] = \frac{\lim_{n\to\infty}(1/n)\sum_{i=1}^n E_{\beta,\xi_i}\left[g_j(x_i, z_i, \beta)^2\right]}{\left\{\lim_{n\to\infty}(1/n)\sum_{i=1}^n E_{\beta,\xi_i}\left[\partial g_j(x_i, z_i, \beta)/\partial\beta_j\right]\right\}^2}, \tag{3.3}$$

where $E_{\beta,\xi_i}[\cdot]$ denotes the expectation with respect to the distributions specified by β and ξ_i, and $\partial g_j(\cdot)/\partial\beta_j$ is the partial derivative with respect to β_j. The derivation of the asymptotic variance (3.3) is given in Appendix A.1.

We extend estimating functions (3.1) and (3.2) to the case with heterogeneous error variances. In order to simplify the estimating equation, we constrained $h(s)$ to a linear function, $h(s) = as + b$. The resulting estimating equations become

$$\sum_{i=1}^n \frac{z_i - \hat{\beta}_0 - \hat{\beta}_1 x_i}{\sigma_{zi}^2 + \hat{\beta}_1^2\sigma_{xi}^2} = 0 \Leftrightarrow \hat{\beta}_0 = \sum_{i=1}^n \frac{z_i - \hat{\beta}_1 x_i}{\sigma_{zi}^2 + \hat{\beta}_1^2\sigma_{xi}^2} \Big/ \sum_{i=1}^n \frac{1}{\sigma_{zi}^2 + \hat{\beta}_1^2\sigma_{xi}^2} \tag{3.4}$$

for β_0 and

$$\sum_{i=1}^n \left\{ a_i \frac{\sigma_{zi}^2 x_i + \hat{\beta}_1\sigma_{xi}^2(z_i - \hat{\beta}_0)}{\sigma_{zi}^2 + \hat{\beta}_1^2\sigma_{xi}^2} + b_i \right\} \frac{z_i - \hat{\beta}_0 - \hat{\beta}_1 x_i}{\sigma_{zi}^2 + \hat{\beta}_1^2\sigma_{xi}^2} = 0 \tag{3.5}$$

for β_1. It is necessary to decide on a_i and b_i to estimate $\hat{\beta}_1$, while $\hat{\beta}_0$ is uniquely determined. If we choose a_i and b_i to minimize the asymptotic variance (3.3), we get $a_i = 0$ and $b_i = \xi_i$. Equation (3.5) then becomes

$$\sum_{i=1}^n \xi_i \frac{z_i - \hat{\beta}_0 - \hat{\beta}_1 x_i}{\sigma_{zi}^2 + \hat{\beta}_1^2\sigma_{xi}^2} = 0. \tag{3.6}$$

$\hat{\beta}_1$ cannot be estimated by Equation (3.6) because it includes the unknown parameter ξ_i. Accordingly, we estimate $\hat{\beta}_1$ by substituting the maximum likelihood estimator $\hat{\xi}_i = \left\{\sigma_{zi}^2 x_i + \hat{\beta}_1 \sigma_{xi}^2 (z_i - \hat{\beta}_0)\right\} / (\sigma_{zi}^2 + \hat{\beta}_1^2 \sigma_{xi}^2)$ of ξ_i into Equation (3.6). This $\hat{\xi}_i$ is consistent with the sufficient statistic s_i for ξ_i. As the result, $a_i = 0$ and $b_i = \xi_i$ derive $h(s_i) = s_i$. The estimating equation for β_1 becomes

$$\sum_{i=1}^{n} \frac{\left\{\sigma_{zi}^2 x_i + \hat{\beta}_1 \sigma_{xi}^2 (z_i - \hat{\beta}_0)\right\} (z_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}{(\sigma_{zi}^2 + \hat{\beta}_1^2 \sigma_{xi}^2)^2}. \tag{3.7}$$

Incidentally, $\hat{\beta}_0$ and $\hat{\beta}_1$ in Equations (3.4) and (3.7) are the same as the maximum likelihood estimators (Walter 1997). The derivations of Equations (3.4) and (3.7) through the maximum likelihood method are given in Appendix A.2. The variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by

$$\begin{aligned} \mathrm{var}(\hat{\beta}_0) &= \frac{\sum f_i \xi_i^2}{\sum f_i \sum f_i \xi_i^2 - \left(\sum f_i \xi_i\right)^2} \\ \mathrm{var}(\hat{\beta}_1) &= \frac{\sum f_i}{\sum f_i \sum f_i \xi_i^2 - \left(\sum f_i \xi_i\right)^2}, \end{aligned}$$

where $f_i = 1/(\sigma_{zi}^2 + \beta_1^2 \sigma_{xi}^2)$ and $\sum = \sum_{i=1}^{n}$.

4. SIMULATION STUDY

We conducted a simulation study to evaluate the performance of the parameter estimation method proposed in Section 3. The simulation study was conducted in a framework that was adaptable to data obtained in a validation study described in Section 2. The following four parameter estimation methods were compared:

1. Ordinary least squares (we call this method OLS).
2. Estimating Equations (3.4) and (3.7) under the assumption of homogeneous error variances, and the error variance of response variables are equal to that of explanatory variables, that is, $\sigma_{xi}^2 = \sigma_{zi}^2 = \sigma^2$ (we call this method EV1).
3. Estimating Equations (3.4) and (3.7) under the assumption of homogeneous error variances, and the error variance of response variables are not equal to that of explanatory variables, that is, $\sigma_{xi}^2 = \sigma_x^2$ and $\sigma_{zi}^2 = \sigma_z^2$ (we call this method EV2).
4. Estimating Equations (3.4) and (3.7) under the assumption of heterogeneous error variances, and all error variances of response variables and explanatory variables vary between measurements (we call this method NEV).

We substituted the Equation (3.4) into Equation (3.7) to find $\hat{\beta}_1$ in EV1, EV2, and NEV. The Newton-Raphson algorithm was used to solve estimating equations. The initial values were set to be OLS estimates. The conditions and steps for the simulation study were as follows:

1. Generate a uniform random number with a range between 5 and 15. This value is set as ξ, which denotes a true measurement in the CAM-TB.

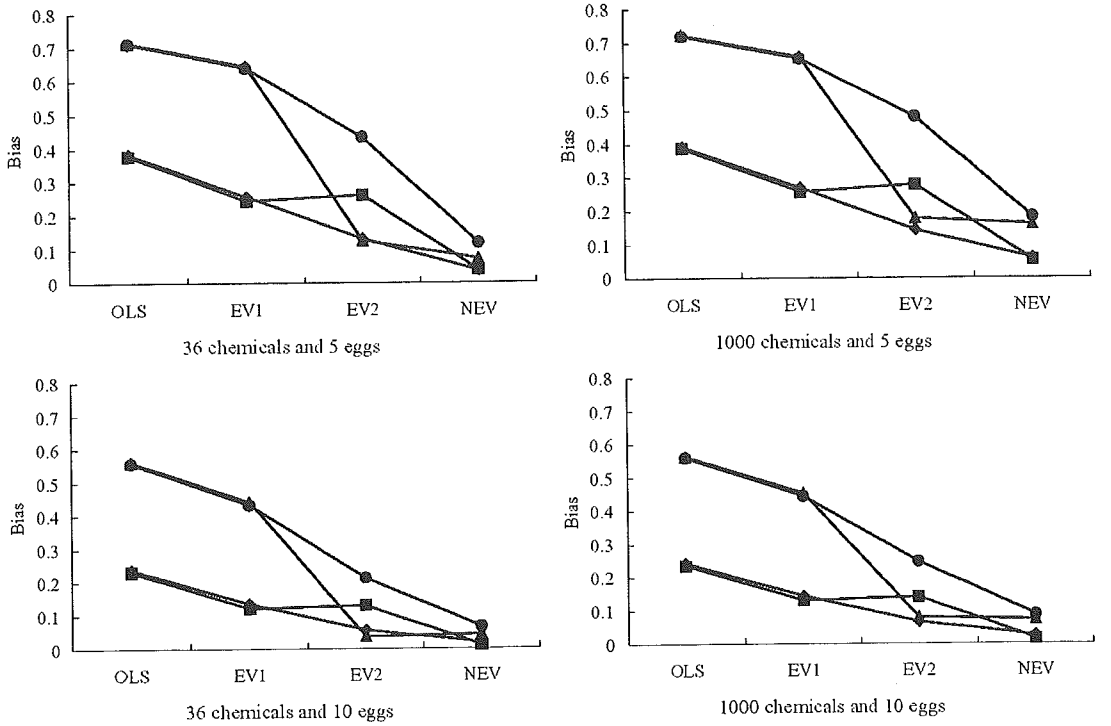

Figure 3. The Results of $\hat{\beta}_1$ for a Simulation Study: Biases (true values minus the means of estimates). Diamonds: $(SD_x, SD_y) = (.5\xi, .0025\,\eta(110-\eta))$, squares: $(SD_x, SD_y) = (.5\xi, .005\,\eta(110-\eta))$, triangles: $(SD_x, SD_y) = (\xi, .0025\,\eta(110-\eta))$, and circles: $(SD_x, SD_y) = (\xi, .005\,\eta(110-\eta))$.

2. Calculate $\eta = 110\exp(\xi - 10)/\{1 + \exp(\xi - 10)\}$ (true model). This η denotes a true score in the Draize test.
3. Generate normal random numbers with mean ξ and standard deviation $\mathrm{SD}_x = .5\,\xi$ or ξ. The number of generated random numbers is 5 or 10, which is assumed to be the number of eggs used in the CAM-TB. Set the mean and standard deviation calculated from these random numbers as a measurement and its according standard deviation in the CAM-TB. Similarly, generate normal random numbers with mean η and standard deviation $\mathrm{SD}_y = .0025\eta(110-\eta)$ or $.005\eta(110-\eta)$. If a generated value is smaller than 0 or larger than 110, the value is replaced with 0 or 110. The number of generated random numbers is 3, which is the same as the number of rabbits used in the Draize test. Set the mean and standard deviation calculated from these random numbers as the score and its according standard deviation in the Draize test.
4. Repeat Steps 1–3 36 times (in correlation to the number of chemical substances) or 1,000 times. This yields one dataset.
5. Estimate parameters (β_0 and β_1) by the four methods described (OLS, EV1, EV2, and NEV).
6. Repeat the above steps 1,000 times, and calculate means, biases (true values minus means), and mean squared errors (MSEs) of parameter estimates ($\hat{\beta}_0$ and $\hat{\beta}_1$).

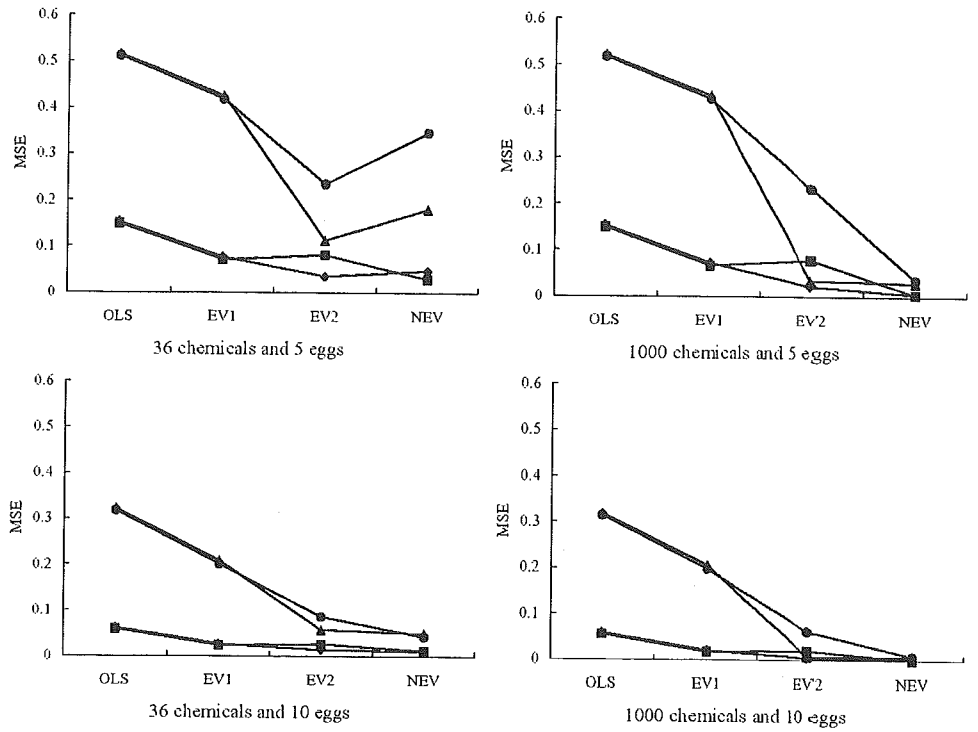

Figure 4. The Results of $\hat{\beta}_1$ for a Simulation Study: Mean squared errors. Diamonds: $(SD_x, SD_y) = (.5\xi, .0025\eta(110 - \eta))$, squares: $(SD_x, SD_y) = (.5\xi, .005\eta(110 - \eta))$, triangles: $(SD_x, SD_y) = (\xi, .0025\eta(110 - \eta))$, and circles: $(SD_x, SD_y) = (\xi, .005\eta(110 - \eta))$.

The results of the simulation study are summarized in Figures 3 and 4. We show only the results of $\hat{\beta}_1$ since $\hat{\beta}_0$ was uniquely determined as a result of $\hat{\beta}_1$. Figure 3 indicates biases on the four methods under $4 \times 2 \times 2 = 16$ conditions (pairs of standard deviations $(\mathrm{SD}_x, \mathrm{SD}_y) = (.5\xi, .0025\eta(110 - \eta))$, $(.5\xi, .005\eta(110 - \eta))$, $(\xi, .0025\eta(110 - \eta))$ or $(\xi, .005\eta(110 - \eta))$; the number of chemical substances = 36 or 1,000; and the number of eggs = 5 or 10). Figure 4 indicates mean squared errors.

The following results were obtained from Figure 3. First, magnitude of biases for the four methods was examined. All four results in Figure 3 show that NEV, EV2, EV1, and OLS generated smaller order biases. However, the biases of EV2 are a little larger than those of EV1 under $(\mathrm{SD}_x, \mathrm{SD}_y) = (.5\xi, .005\eta(110 - \eta))$ (symbols in Figure 3 = squares). The biases of EV2 are also similar to those of NEV under $(\mathrm{SD}_x, \mathrm{SD}_y) = (\xi, .0025\eta(110 - \eta))$ (symbols in Figure 3 = triangles).

Second, we examined how the difference of standard deviations (SD_x or SD_y) influenced biases. All four results in Figure 3 show that biases for $\mathrm{SD}_x = \xi$ (symbols in Figure 3 = triangles and circles) are larger than those for $\mathrm{SD}_x = .5\xi$ (symbols in Figure 3 = diamonds and squares). On the other hand, the difference of SD_y, that is, $\mathrm{SD}_y = .0025\eta(110 - \eta)$ (symbols in Figure 3 = diamonds and triangles) or $\mathrm{SD}_y = .005\eta(110 - \eta)$ (symbols in Figure 3 = squares and circles), does not generate much difference in biases except for EV2. These results imply that error variances in the CAM-TB have more influence on biases than those of the Draize test. This is especially true for OLS and EV1. It seems that

Table 1. Parameter Estimates and Their Standard Errors by Four Estimation Methods (OLS: ordinary least squares; EV1: $\sigma_{xi}^2 = \sigma_{zi}^2 = \sigma^2$; EV2: $\sigma_{xi}^2 = \sigma_x^2$ and $\sigma_{zi}^2 = \sigma_z^2$ but $\sigma_x^2 \neq \sigma_z^2$; and NEV: all error variances vary between measurements)

	$\hat{\beta}_0$	s.e. $(\hat{\beta}_0)$	$\hat{\beta}_1$	s.e. $(\hat{\beta}_1)$
OLS	−6.189	.673	.277	.047
EV1	−6.399	1.393	.298	.098
EV2	−6.281	1.633	.287	.114
NEV	−10.619	.380	.864	.150

error variances in the Draize test have no relevance to biases. However, the biases of EV2 are small when error variances in the CAM-TB are large.

Third, we examined how differences in the numbers of chemical substances or the number of eggs influenced biases. The comparisons between the same numbers of chemical substances and different numbers of eggs (the comparisons between top and bottom in Figure 3) show that on all of four methods, biases when the number of eggs = 5 are larger than those for when the number of eggs = 10. On the other hand, the comparisons between the different numbers of chemical substances and same numbers of eggs (the comparisons between left and right in Figure 3) show that the magnitude of bias does not change with changes in the number of chemical substances. These results imply that biases become smaller if the number of eggs used in the CAM-TB increases but not smaller even if the number of chemical substances increases.

In Figure 4, we can see that MSEs also have features similar to biases. However, NEV produces large MSEs in the case in which the number of chemical substances is 36 and the number of eggs is 5.

5. APPLICATION OF THE PROPOSED METHOD TO REAL DATA

We applied the proposed method and the other 3 methods to the data described in Section 2.

Table 1 summarizes the regression parameters and their standard errors estimated by the four methods described in Section 3. Figure 5 shows their fitted curves. $\hat{\beta}_0$ and $\hat{\beta}_1$ by OLS, EV1, and EV2 (solid line, large broken line, and small broken line in Figure 5) provided almost identical results. Those provided by NEV (solid thin line in Figure 5) were considerably differed from the other three methods. As Figure 5 indicates, the four methods do not fit the data well.

Our objective is to formulate a model to predict the score in the Draize test by the measurement in the CAM-TB. Accurate prediction of Draize test scores is much more important when CAM-TB produces moderate measurements. Large CAM-TB measurements indicate high Draize test scores, even though the prediction is not very precise. Though the prediction is loose, it is possible to identify chemical substances that have strong toxicity. A chemical substance having strong toxicity cannot be adopted for use in cosmetics. When the CAM-TB test produces moderate measurements, however, we should be careful to avoid underestimating the toxicity of the chemical substance being tested. This is what happens

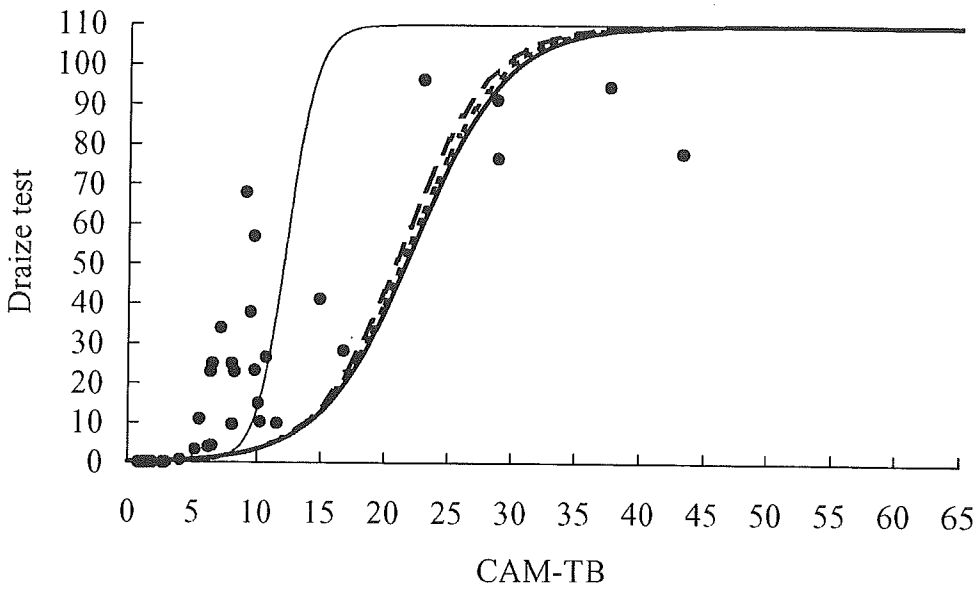

Figure 5. Fitted curves (solid line: OLS, large broken line: EV1, small broken line: EV2, and solid thin line: NEV).

when the score in the Draize test is underestimated.

Figure 5 shows that OLS, EV1, and EV2 underestimate the score in the Draize test when the CAM-TB test produces moderate measurements. Figure 5 also shows that NEV makes the degree of underestimation smaller in the moderate measurements in comparison with the other three methods. NEV is a worse fit than the other three methods for large measurements. As discussed earlier, however, it is not important that the fit for these measurements is wrong. Consequently, we consider that the proposed method, NEV, results in a better prediction formula than the other three methods.

6. DISCUSSION

In this article, we extended the estimating function method proposed by Amari and Kawanabe (1997) to the case in which error variances vary between measurements. The estimators for the proposed method were incidentally identical to the maximum likelihood estimators.

We also conducted a simulation study to compare the regression parameters estimated by the proposed method with those estimated by three other methods under the assumption of homogeneous error variances. The parameters estimated by the proposed method, NEV, may be imprecise when sample size is small and error variances are large. However, the results of the simulation study suggest that, in comparison to the other three methods, NEV results in smaller biases and MSEs. As shown in Figure 4, NEV produced large MSE in the case in which the number of chemical substances was 36 and the number of eggs was 5. The reason is that there are few cases in which NEV takes a very large value of $\hat{\beta}_1$ when the error variances are large. For example, under $(\mathrm{SD}_x, \mathrm{SD}_y) = (\xi, .005\eta(110 - \eta))$, NEV took 16.93 as the maximum value of $\hat{\beta}_1$.

In the simulation study, normal random numbers were generated on the original scales before the logit transformation, while the scales after the logit transformation were modeled. Our primary objective was to see if the proposed method could have worked in the validation study. In the validation study, measurement errors occurred at the original scale. This was because we considered normal error structures at the original scale in the simulation study. If we knew the true error structures, it would be appropriate to use a nonlinear measurement error model. Note that to develop such a nonlinear measurement error model is troublesome with heterogeneous error variances. As a result, in the simulation study, we used somewhat misspecified models. Nevertheless, the simulation study showed that the proposed method, NEV, gave the smaller biases and MSEs than the other three methods.

The goal of the validation study was to predict Draize test scores from CAM-TB values. When a fitted curve shows greater deviation in the direction of underestimating the score in the Draize test, there is a greater risk of underestimating a chemical substance's toxicity. From the results of the simulation study shown in Section 3, it is reasonable to conclude that the proposed method is less biased than the other three methods in the validation study. The results of the simulation study also indicate that we can make bias and MSE smaller if larger numbers of eggs are used in the CAM-TB. This shows that if alternative methods such as the CAM-TB are carried out with large numbers of experiment materials for each chemical substance, we can estimate the regression parameters more precisely.

In validation studies, both the measurements in animal experiments and in alternative methods are measured with error. We need to adopt measurement error models when we predict measurements in animal experiments by alternative methods. In order to investigate whether or not the method proposed in this paper is useful in evaluating alternative methods, we need to examine other alternative methods in addition to CAM-TB.

APPENDIX

A.1 THE DERIVATION OF ASYMPTOTIC VARIANCE (3.3)

If the estimating equation for $\hat{\beta}_j (j = 0, 1)$, $\sum_{i=1}^{n} g_j(x_i, z_i, \hat{\beta}_j, \beta_k) = 0\,(j \neq k)$, is expanded with a Taylor series on the true parameter β_j, we get

$$0 = \sum_{i=1}^{n} g\left(x_i, z_i, \beta\right) + \sum_{i=1}^{n} \frac{\partial g(x_i, z_i, \beta)}{\partial \beta_j} \left(\hat{\beta}_j - \beta_j\right) + O_p\left(\left|\hat{\beta}_j - \beta_j\right|^2\right). \tag{A.1}$$

$\hat{\beta}_j$ takes the almost same value as β_j when n becomes very large. Then, since the higher order terms, $O_p\left(\left|\hat{\beta}_j - \beta_j\right|^2\right)$, are negligible, Equation (A.1) is transformed as

$$\sqrt{n}\left(\hat{\beta}_j - \beta_j\right) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(x_i, z_i, \beta\right) \Bigg/ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g\left(x_i, z_i, \beta\right)}{\partial \beta_j}. \tag{A.2}$$

By applying the law of large numbers, the denominator on the right hand side in Equation (A.2) converges to

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E_{\beta,\xi_i} \left[\frac{\partial g_j(x_i, z_i, \beta)}{\partial \beta_j} \right].$$

By the central limit theorem, the numerator on the right hand side in Equation (A.2) is distributed according to

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_j(x_i, z_i, \beta) \sim N\left(0, E_{\beta,\xi_i}\left[g_j(x_i, z_i, \beta)^2\right]\right).$$

Therefore, $\hat{\beta}_j$ is also asymptotically normally distributed with the asymptotic variance

$$\text{a.v.}\left[\sqrt{n}\left(\hat{\beta}_j - \beta_j\right)\right] = \frac{\lim_{n\to\infty}(1/n)\sum_{i=1}^{n} E_{\beta,\xi_i}\left[g_j(x_i, z_i, \beta)^2\right]}{\left\{\lim_{n\to\infty}(1/n)\sum_{i=1}^{n} E_{\beta,\xi_i}\left[\partial g_j(x_i, z_i, \beta)/\partial\beta_j\right]\right\}^2}.$$

A.2 THE DERIVATIONS OF EQUATIONS (3.4) AND (3.7) THROUGH THE MAXIMUM LIKELIHOOD METHOD

A pair of measurements, (x_i, z_i), $i = 1, 2, \ldots, n$, have a functional relationship

$$\begin{aligned} x_i &= \xi_i + \delta_i \\ z_i &= \beta_0 + \beta_1 \xi_i + \varepsilon_i, \end{aligned}$$

where δ_i and ε_i are distributed according to

$$\begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{xi}^2 & 0 \\ 0 & \sigma_{zi}^2 \end{pmatrix}\right).$$

The log-likelihood function, $\log L$, is given as

$$\log L = C - \frac{1}{2} \sum_{i=1}^{n} \left\{ \frac{(x_i - \xi_i)^2}{\sigma_{xi}^2} + \frac{(z_i - \beta_0 - \beta_1 \xi_i)^2}{\sigma_{zi}^2} \right\},$$

where C denotes a set of constant terms that does not depend on the parameters β_0, β_1, and ξ_i. $\partial \log L/\partial \xi_i = 0$ gives the maximum likelihood estimator

$$\hat{\xi}_i = \frac{\sigma_{zi}^2 x_i + \hat{\beta}_1 \sigma_{xi}^2 (z_i - \hat{\beta}_0)}{\sigma_{zi}^2 + \hat{\beta}_1^2 \sigma_{xi}^2}.$$

By substituting this $\hat{\xi}_i$ into $\partial \log L/\partial \beta_0 = 0$ and $\partial \log L/\partial \beta_1 = 0$, Equations (3.4) and (3.7) are derived.

ACKNOWLEDGMENTS

The authors thank Takashi Omori and the referees for their helpful comments.

[Received February 2004. Revised May 2004.]

REFERENCES

Amari, S., and Kawanabe, M. (1997), "Information Geometry of Estimating Functions in Semi-Parametric Statistical Models," *Bernoulli*, 3, 29–54.

Draize, J., Woodard, G., and Galvery, H. (1944), "Methods for the Study of Irritation and Toxicity of Substances Applied Topically to the Skin and Mucous Membranes," *The Journal of Pharmacology and Experimental Therapeutics*, 82, 377–390.

Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley.

Hagino, S., Itagaki, H., Kato, S., and Kobayashi, T. (1993), "Further Evaluation of the Quantitative Chorioallantoic Membrane Test Using Trypan Blue Stain to Predict the Eye Irritancy of Chemicals," *Toxicology in Vitro*, 7, 35–39.

Hagino, S., Itagaki, H., Kato, S., Kobayashi, T., and Tanaka, M. (1991), "Quantitative Evaluation to Predict the Eye Irritancy of Chemicals Modification of Chorioallantoic Membrane Test by Using Trypan Blue," *Toxicology in Vitro*, 5, 301–304.

Hagino, S., Kinoshita, S., Tani, N., Nakamura, T., Ono, N., Konishi, K., Iimura, H., Kojima, H., and Ohno, Y. (1999), "Interlaboratory Validation of In Vitro Eye Irritation Tests for Cosmetic Ingredients (2) Chorioallantoic Membrane (CAM) Test," *Toxicology in Vitro*, 13, 99–113.

Japan Cosmetic Industry Association (1994), *Report of the First Validation Study on Alternative Methods to the Draize Eye Irritation Test*, Japan Cosmetic Industry Association (in Japanese).

Neyman, J., and Scott, E. L. (1948), "Consistent Estimates Based on Partly Consistent Observation," *Econometrica*, 32, 1–32.

OECD Test Guidelines Programme (1996), "Options for a Testing Strategy for the Testing of Skin and Eye Irritancy," Final report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods, September 18–19, 1996, the Chateau de la Muette, Paris.

Walter, S. D. (1997), "Variation in Baseline Risk as an Explanation of Heterogeneity in Meta-analysis," *Statistics in Medicine*, 16, 2883–2900.

Wilhelmus, K. R. (2001), "The Draize Eye Test," *Survey of Ophthalmology*, 45, 493–515.

CLINICAL TRIALS ARTICLE Clinical Trials 2005; 2: 174–181

Overlap coefficient for assessing the similarity of pharmacokinetic data between ethnically different populations

Sachiko Mizuno, Takuhiro Yamaguchi, Akira Fukushima, Yutaka Matsuyama and Yasuo Ohashi

We developed a method to assess the similarity of pharmacokinetic data between ethnically different populations. An evaluation of confidence intervals for the mean difference in pharmacokinetic parameters, such as area under the concentration-versus-time curve (AUC), between populations is often used. We propose the use of the overlap coefficient (OC), which represents the proportion of overlap between two probability distributions, as a measure of the similarity between distributions. We considered five OC estimators – two parametric ones and three nonparametric ones. Simulation studies were conducted to compare the performance of the five OC estimators and their bootstrap confidence intervals. Results showed that nonparametric estimators with fixed-bandwidth kernel density estimation had a smaller mean squared error in almost all situations, and their coverage probabilities were close to the nominal level. The proposed method was applied to pharmacokinetic data from a bridging study of a combination therapy for metastatic colorectal cancer patients in the USA and Japan. From the analyses of this study, it was suggested that the distributions of the logarithmically transformed AUC for leucovorin and 5-fluorouracil were similar between the two populations. *Clinical Trials* 2005; **2**: 174–181. www.SCTjournal.com

1 Introduction

In the development of new medical products, clinical trials are required to be standardized and conducted efficiently and quickly on a global scale. In an effort to address these issues, the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) was initiated in 1990. One outcome of the project was the development of the guideline on "Ethnic Factors in the Acceptability of Foreign Clinical Data" [1], known as the ICH E5 guideline. One of the concerns of E5 is a bridging study to allow extrapolation of foreign clinical data to a new region. The E5 guidance is based on the premise that it is not necessary to repeat the entire clinical drug development programme in the new region; bridging studies should allow for new medicines to be supplied expeditiously to patients for their benefit.

When evaluating the extrapolation of clinical data from one region to an ethnically different region, it is important to assess whether the pharmacokinetic data are similar across the populations. Pharmacokinetic studies are conducted to characterize the absorption, distribution, metabolism, and excretion of a drug either in blood or in other pertinent locations. The pharmacokinetic profiles of two populations are compared in terms of the appropriate parameters, which are measures of systemic exposure such as peak concentration and area under the concentration-versus-time curve (AUC). However, there are no standard methods for assessing the similarity of pharmacokinetic parameters. Graphical presentation of data and the evaluation of confidence intervals for the difference

Department of Biostatistics, School of Health Sciences and Nursing, University of Tokyo, Tokyo, Japan
Author for correspondence: Sachiko Mizuno, Department of Biostatistics, School of Health Sciences and Nursing, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: mizuno@epistat.m.u-tokyo.ac.jp

10.1191/1740774505cn077oa

14. **Spiegelhalter DJ, Best NG, Carlin BP, Linde AVD.** Bayesian measures of complexity and fit. *Journal of the Royal Statistical Society Series B* 2003; **64**: 583–639.
15. **Wallace P, Haines A, Harrison R** *et al.* Joint teleconsultations (virtual outreach) versus standard outpatient appointments for patients referred by their general practitioner for a specialist opinion: a randomised trial. *The Lancet* 2002; **359**: 1961–68.
16. **Ware JE, Snyder MK, Wright WR, Davies AR.** Defining and measuring patient satisfaction with medical care. *Eval Program Plann* 1983; **6**: 247–63.
17. **Brealey S, Burton K, Coulton S** *et al.* UK back pain exercise and manipulation (UK BEAM) trial—national randomised trial of physical treatments for back pain in primary care: objectives, design and interventions. *BMC Health Serv Res* 2004; **3**: 16.
18. **Roland M, Morris RW.** A study of the natural history of back pain: 1 – development of a reliable and sensitive measure of disability in low back pain. *Spine* 1983; **8**: 141–44.
19. **Senn SJ.** Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* 1989; **8**: 467–75.
20. **Ribaudo HJ, Thompson SG.** The analysis of repeated multivariate binary quality of life data: a hierarchical model approach. *Statistical Methods in Medical Research* 2002; **11**: 69–83.
21. **Hoover DR.** Clinical trials of behavioural interventions with heterogeneous teaching subgroup effects. *Statistics in Medicine* 2002; **21**: 1351–64.
22. **Butler C, Bachmann M.** Design and analysis of studies evaluating smoking cessation interventions where effects vary between practices and practitioners. *Family Practice* 1996; **13**: 402–407.
23. **Morris HF, Ocki S, Winkler S.** Implant survival in patients with type 2 diabetes: placement to 36 months. *Annals of Periodontology* 2000; **5**: 157–65.
24. **Williams AC, Sandy JR, Thomas S, Sell D, Sterne JAC.** Influence of a surgeon's experience on speach outcome in cleft lip and palate. *The Lancet* 1999; **354**: 1697–98.
25. **Mor V, Laliberte LL, Petrisek AC** *et al.* Impact of breast cancer treatment guidelines on surgical practice patterns: results of a hospital-based intervention. *Surgery* 2000; **128**: 847–61.
26. **Elkin I.** A major dilemma in psychotherapy outcome research: disentangling therapists from therapies. *Clin Psychol Sci Prac* 1999; **6**: 10–32.
27. **Jones B, Teather D, Wang J, Lewis JA.** A comparison of various estimators of a treatment difference for a multi-centre clinical trial. *Statistics in Medicine* 1998; **17**: 1767–77.
28. **Natarajan R, Kass RE.** Reference Bayesian methods for generalised linear mixed models. *Journal of American Statistical Association* 2000; **95**: 227–37.
29. MLwiN: http://multilevel.ioe.ac.uk/index.html 2003 (accessed 4 February 2005).
30. **Liang KY, Zeger SL.** Longitudinal data analysis using generalised linear models. *Biometrika* 1986; **73**: 13–22.

33