Determination of Reference Intervals for 13 Plasma Proteins Based on IFCC International Reference Preparation (CRM470) and NCCLS Proposed Guideline (C28-P, 1992): A Strategy for Partitioning Reference Individuals With Validation Based on Multivariate Analysis

Kiyoshi Ichihara^{1*} and Tadashi Kawai²

¹Department of Clinical Pathology, Kawasaki Medical School, Kurashiki, Japan

²Department of Clinical Pathology, Jichi Medical School, Tochigi, Japan

In our recent project of establishing reference intervals for 13 plasma proteins following the NCCLS proposed guideline (C28-P, 1992), confounding and interacting relationships were found among criteria for partitioning reference individuals. Therefore, we found it necessary to test for the validity of the criterion by multivariate analysis so that such relationships could be revealed. Meanwhile, to find difference appropriate for the partition, sex- and age-related differences were examined in common screening tests whose reference intervals (RI) were conventionally set for each subgroup. The result led to our standard for partitioning when the difference between subgroups exceed 10% of the span of a RI derived without partitioning. Among the 13 plasma proteins, the multi-

variate analyses and the standard indicated gender-based RI was for IgM, α_1 acid glycoprotein, α_2 macroglobulin, and transferrin, transthyretin. None of the age-related differences were large enough. In serum IgM, ceruloplasmin, and transferrin, however, females tended to have larger differences between ages, exemplifying interaction of gender on age-related changes of the analytes. Smoking-related differences were > 10% in IgG and transthyretin. However, differences in the latter proved to be spurious multivariately. These results suggest a necessity for additional descriptions on the validity of the partitioning criterion in the NCCLS proposed guideline. *J. Clin. Lab. Anal.* 11:117–124.

© 1997 Wiley-Liss, Inc.

Key words: partitioning criterion; multiple regression; multivariate analysis; multiple logistic regression; confounding; interaction; sex- and age-related differences; smoking

INTRODUCTION

The guideline “How to define, determine, and utilize reference intervals in clinical laboratory,” proposed by the National Committee for Clinical Laboratory Standards (NCCLS) (1), notes that partitioning of a reference sample should be considered when there is a statistically significant difference among subgroups specified by such factors as age, sex, and race, including smoking and drinking habits. The guideline, however, did not make note of the possibility of hidden relationships among those factors.

In fact, in our current project of establishing reference intervals for 13 plasma proteins (2), we noticed a significantly lower serum IgM concentration in smokers than in nonsmokers. Actually, it turned out to be a spurious relationship attributable to a sex difference in IgM concentration (lower IgM in male), since the majority of the smokers were males and there was a higher proportion of females among the nonsmokers (Fig. 1). In statistical terms, such a phenomenon is called “con-

founding” of one factor (gender) in the analysis of an effect of another (smoking) on a particular test value (IgM concentration) (3, 4).

Similarly, we experienced that, although there was no apparent age-related difference in IgM concentration, when it was analyzed separately for each sex, we observed the age difference in IgM concentration in females but not in males (Fig. 2). In statistical terms, such a phenomenon is called “interaction” of one factor (gender) in the analysis of an effect of another (age) on a particular test value (IgM concentration) (3, 4).

These situations clearly indicated that simple subgrouping

*Correspondence to: Kiyoshi Ichihara, M.D., Department of Clinical Pathology, Kawasaki Medical School, 577 Matsushima, Kurashiki, 701-01 Japan.

Received 6 August 1996; Accepted 7 August 1996

Fig. 1. Confounding of sex difference on the relationship between smoking habit and serum IgM. (a) Difference in IgM between smokers and nonsmokers. (b) Difference in IgM between males and females. (c) Differences

in IgM concentrations between smokers and nonsmokers for each sex, demonstrating that the lower serum IgM in smokers shown in (a) is spurious and is attributable to the sex difference.

Fig. 2. Interaction of sex on the relationship between age and serum IgM. Top panel shows an age-related change in serum IgM analyzed from the 422 reference individuals. The subjects were divided into four age subgroups, up to 39, 40–49, 50–59, and > 60 years. Middle and bottom panels correspond to the relationship for males and females, respectively, demonstrating an age-related change in females that is not apparent in the top panel. This situation is referred to as an “interaction” of sex on the relationship between age and serum IgM.

of reference individuals based on a single partitioning criterion can be quite misleading. Consequently some multivariate approach will be essential to adjust for possible confounding relationships or interactions among factors and to identify a factor truly relevant for the subgrouping.

In this report, we present two multivariate procedures to test for the validity of a partitioning criterion: multiple regression and multiple logistic regression analyses (3, 5). Herein, we also introduce another guideline to determine whether an observed difference among subgroups specified by a partitioning criterion is large enough to subdivide a given reference sample in order to establish separate reference intervals.

MATERIALS AND METHODS

Subjects and Data Served for Analysis

The subjects and data we used for the analysis were identical to those in our previous report (2). A brief outline follows.

Among adults who applied for a 2-day health check program at Kanagawa Public Health Clinic during a 3-month period, 999 healthy adults (27–77) years of age were chosen as subjects for reference individuals by medical interviewing, physical examination, and a questionnaire regarding health status.

Blood samples were obtained from them in the morning at a fasting condition and tested for the following 25 items on the day of the sampling: creatinine (Crn), urea nitrogen (UN), uric acid (UA), Na, K, fasting blood sugar (fBS), hemoglobin A_{1c} (HbA_{1c}), aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma glutamyl transferase (GGT), alkaline phosphatase (ALP), thymol turbidity test (ZTT), lactic dehydrogenase (LDH), choline esterase (ChE), total protein (TP), albumin (Alb), white blood cell (WBC), red blood cell (RBC), hemoglobin (Hb), hematocrit (Ht), platelet (PLT), total cholesterol (TCH), high density lipoprotein-cholesterol (HDL), triglyceride (TG), lipid peroxide (LPO). The serum concentration of Alb was calculated as a fraction of albumin determined by serum electrophoresis multiplied by the concentration of TP.

The rest of the serum specimens were stored at –70°C for a couple of months and further tested for 13 plasma proteins at seven laboratories (Jichi Medical School, Hamamatsu Medical School, Nihon University, Tokyo University, Special Reference Laboratory, Hoechst Japan): immunoglobulins G, A, M (IgG, IgA, IgM), complement components C3 and C4 (C3, C4), transferrin (Tf), α_1 antitrypsin (α_1 AT), α_1 acid glycoprotein (α_1 AG), haptoglobin (Hp), α_2 macroglobulin (α_2 M), celuroplasmin (Cp), albumin (Alb), transthyretin or prealbumin (TTR) by use of nephelometry, Behring Nephelometric Analyzer (BNA), employing the IFCC international reference preparation for plasma proteins, CRM470, as its calibration standard (6).

During the study period, the intra-assay and interassay

variations, monitored using an identical external quality assurance program at the seven laboratories, were comparable and all within an acceptable range. Thus the results were pooled to calculate the reference intervals for the 13 plasma proteins.

Analyses of the Validity of a Partitioning Criterion

Analysis for variations among subgroups

The variations among subgroups specified by a given partitioning criterion were calculated by the principle of oneway analysis of variance (ANOVA) (7). Here the variations were expressed not as a variance but as a standard error. A pure fraction of between-subgroup variations s_G , subtracted of a component of within-subgroup variations s_I , was calculated by the following formulae:

$$s_{G_0} = \frac{\sum_{i=1}^k n_i (\bar{x}_i - \bar{\bar{x}})^2}{k-1}$$

$$s_I = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} n_i (x_{ij} - \bar{x}_i)^2}{\sum_{i=1}^k (n_i - 1)}$$

$$s_G = \sqrt{s_{G_0}^2 - \frac{s_I^2}{k}}$$

where k is the number of subgroups, n_i is the number of subjects in the i -th subgroup, $\bar{x}_i$, the mean for the i -th subgroup, $\bar{\bar{x}}$, the grand mean, and s_{G_0} represents gross between-subgroup variations.

The pure fraction of between-subgroup variations was standardized by dividing it with the span of the reference interval (RI), between the lower and upper limits, which was derived from the entire data without subgrouping.

between subgroup variations relative to RI = $\frac{s_G}{\text{span of RI}} \cdot 100$

Multivariate analysis

Multiple regression analysis. Significance of a given criterion to subgroup the reference values was evaluated by the multiple regression analysis so that potential effects of other factors could be statistically adjusted. The following two models were introduced according to the principle presented by Kleinbaum DG (3) to reveal factors (x_2 - x_p), which

may potentially get confounded or interact with a given factor of interest (x_1).

A model without interaction terms:

$$y = a_0 + \underbrace{a_1 x_1}_{\text{criterion tested}} + \underbrace{a_2 x_2 + a_3 x_3 + \dots + a_p x_p}_{\text{possible confounding factors}}$$

y : dependent variable (value of a test to determine the reference interval)

x_1 : a variable of interest (a partitioning criterion to be validated)

x_{2-p} : other variables to adjust for possible confounding influences

A model with interaction terms:

$$y = a_0 + \underbrace{a_1 x_1}_{\text{criterion tested}} + \underbrace{a_2 x_2 + a_3 x_3 + \dots + a_p x_p}_{\text{possible confounding factors}} + \underbrace{b_1 x_1 x_2 + b_2 x_1 x_3 + \dots + b_{p-1} x_1 x_p}_{\text{possible interaction}}$$

y : a dependent variable (value of a test to determine the reference interval)

x_1 : a variable of interest (a partitioning criterion to be validated)

x_{2-p} : other variables to adjust for possible confounding influences

$x_1 x_2, x_1 x_3, \dots, x_1 x_p$: terms to adjust for interaction of x_2 through x_p on x_1

Multiple logistic regression. When a partitioning criterion was binomial, its validity was also evaluated in a reverse way by the multiple logistic regression analysis (3), with the value of the criterion (either 0 or 1) set as a dependent variable and other factors or test results as explanatory variables.

$$p = \frac{1}{\exp - (a_0 + a_1 x_1 + a_2 x_2 + \dots + a_p x_p)}$$

where p corresponds to a probability of someone with a set of explanatory variables (x_1 through x_p) having a characteristic specified by the criterion.

RESULTS

Age- and Sex-related Changes in Biochemical Screening Tests

The magnitude of sex-related or age-related changes relative to the span of a reference interval (RI) were calculated from the 25 common screening tests as shown in Table 1. Although there were statistically significant age-related changes in ChE and TCH, their %RIs were < 7%. Actually, their reference intervals are not usually partitioned by age. Meanwhile, sex-related changes in Crn, UA, HDL, RBC, Hb, and Ht, whose RIs are conventionally set separately for each sex, were statistically significant and all the magnitude were > 10% of the span of the RIs.

TABLE 1. Age-, Sex- and Smoking-related Changes in 25 Common Screening Tests^a

Item	Unit	Age			Sex			Smoking		
		s _G	%RI	P	s _G	%RI	P	s _G	%RI	P
Crn	mg/dl	0.01	1.1%	0.24	0.17	19.3%	0.00	0.07	8.1%	0.00
UN	mg/dl	0.56	4.2%	0.01	0.49	3.7%	0.02	0.00	0.0%	0.65
UA	mg/dl	0.34	6.4%	0.02	0.99	18.7%	0.00	0.53	10.1%	0.00
Na	mmol/l	0.30	3.8%	0.00	0.56	7.0%	0.00	0.30	3.8%	0.00
K	mmol/l	0.00	0.0%	0.69	0.05	4.5%	0.01	0.04	4.2%	0.01
fBS	mg/dl	1.19	3.5%	0.02	2.35	6.9%	0.00	0.00	0.0%	0.89
A _{1c}	%	0.10	7.1%	0.00	0.02	1.3%	0.19	0.03	2.3%	0.07
AST	IU/l	0.69	3.7%	0.01	0.65	3.5%	0.01	0.20	1.1%	0.22
ALT	IU/l	0.00	0.0%	0.75	2.86	9.6%	0.00	1.34	4.5%	0.00
γGT	IU/l	1.39	2.0%	0.21	10.39	15.0%	0.00	8.91	12.9%	0.00
ALP	IU/l	8.11	5.7%	0.00	0.00	0.0%	0.36	0.00	0.0%	0.52
ZTT	Kunkel	0.34	3.5%	0.02	0.79	8.1%	0.00	0.71	7.3%	0.00
LDH	IU/l	14.25	7.8%	0.00	13.52	7.4%	0.00	9.82	5.4%	0.00
ChE	ΔpH	0.01	1.8%	0.19	0.05	7.8%	0.00	0.02	2.4%	0.07
TP	g/dl	0.02	1.5%	0.18	0.02	1.4%	0.14	0.09	6.0%	0.00
Alb	g/dl	0.07	6.8%	0.00	0.08	7.9%	0.00	0.00	0.3%	0.30
WBC	/μl	0.00	0.0%	0.49	480.08	8.1%	0.00	755.12	12.7%	0.00
RBC	10 ⁴ /μl	7.16	4.5%	0.00	30.66	19.4%	0.00	11.84	7.5%	0.00
Hb	g/dl	0.15	2.5%	0.06	1.34	23.1%	0.00	0.69	11.9%	0.00
Ht	%	0.38	2.5%	0.09	3.92	25.5%	0.00	2.11	13.7%	0.00
PLT	10 ⁴ /μl	0.33	1.6%	0.21	0.00	0.0%	0.87	0.30	1.5%	0.17
TCH	mg/dl	12.12	8.6%	0.00	3.92	2.8%	0.03	5.17	3.7%	0.01
HDL	mg/dl	0.00	0.0%	0.82	8.94	13.6%	0.00	5.41	8.2%	0.00
TG	mg/dl	14.60	6.6%	0.00	25.04	11.3%	0.00	16.48	7.4%	0.00
LPO	mmol/ml	0.10	4.9%	0.00	0.09	4.3%	0.00	0.07	3.6%	0.01

^aAfter partition of the reference sample by Age, Sex, or Smoking, the between-subgroup variations s_G, its ratio to the reference interval (%RI), which were obtained without partition, and its statistical significance in terms of probability by one way analysis of variance were derived and tabulated for 25 common screening tests.

From this preliminary analysis, we decided to separate the reference samples into subgroups when the between-subgroup variations exceeds 10% of the span of RI.

Age-, Sex- and Smoking-related Changes in 13 Plasma Proteins

Table 2 shows the variations between ages and between sexes relative to RIs for the 13 plasma proteins. Age-related variations were all within 10% of the RI. Sex-related variations were > 10% in IgM, α₁AG, α₂M, and Tf. The smoking-related variations were > 10% in IgG and TTR.

Validation of Partitioning Criterion by Multiple Regression Analysis

The validity of age (AGE) as a partitioning criterion for serum IgM, e.g., was assessed by using the following regression model without an interaction term with the gender (SEX) and habit of smoking (SMK) as terms to check confounding.

$$\text{IgM} = a_0 + a_1(\text{AGE}) + a_2(\text{SEX}) + a_3(\text{SMK})$$

where the SEX and SMK were considered as binomial variables (0/1) and AGE as an integer variable. SEX showed the highest, and AGE showed a marginal association with the

IgM level, whereas SMK showed no significant association with the IgM concentration (Table 3.1). These results are only partially consistent with the one obtained in the univariate analysis shown in Table 2. It is noteworthy that SMK did not actually have an independent relationship with the IgM concentration as revealed in the two-layers stratified analysis shown in Figure 1.

Meanwhile, if another model with an interaction term,

$$\text{IgM} = a_0 + a_1(\text{AGE}) + a_2(\text{SEX}) + a_3(\text{SMK}) + b_1(\text{AGE} \times \text{SEX})$$

was applied, AGE and AGE × SEX turned out to be closely associated with IgM concentration (Table 3.1) as demonstrated in Figure 2. This result indicates the importance of evaluating “interaction” among parameters.

Similarly, the association of SMK on IgG concentration was analyzed using the following formula:

$$\text{IgG} = a_0 + a_1(\text{SMK}) + a_2(\text{AGE}) + a_3(\text{SEX})$$

As shown in Table 3.2, serum IgG was significantly associated with SMK even if analyzed collectively together with AGE and SEX. It is noteworthy that a moderate sex difference (%RI = 8.7%) observed in the univariate analysis (Table 2) turned out to be only marginal, clearly indicating a confounding influence of SMK in the relationship between SEX and serum IgG.

TABLE 2. Age-, Sex- and Smoking-related Changes in 13 Plasma Proteins^a

Item	Unit	Age			Sex			Smoking		
		S _G	%RI	P	S _G	%RI	P	S _G	%RI	P
IgG	mg/dl	28.99	3.2%	0.05	79.69	8.7%	0.00	92.60	10.1%	0.00
IgA	mg/dl	10.73	3.3%	0.04	12.53	3.8%	0.01	0.00	0.0%	0.79
IgM	mg/dl	2.38	1.2%	0.31	24.35	12.0%	0.00	14.96	7.4%	0.00
C3	mg/dl	1.81	2.1%	0.17	0.00	0.0%	0.35	0.00	0.0%	0.60
C4	mg/dl	1.24	3.9%	0.02	1.02	3.2%	0.04	0.72	2.3%	0.11
Tf	mg/dl	5.77	4.0%	0.05	13.43	9.4%	0.00	7.99	5.6%	0.01
α1AT	mg/dl	1.72	2.8%	0.20	0.00	0.0%	0.92	1.47	2.4%	0.17
α1AG	mg/dl	0.00	0.0%	0.99	6.45	11.5%	0.00	4.51	8.0%	0.00
Hp	mg/dl	5.19	3.1%	0.09	0.00	0.0%	0.59	10.29	6.2%	0.00
α2M	mg/dl	7.60	5.8%	0.01	23.42	17.7%	0.00	12.69	9.6%	0.00
Cp	mg/dl	1.36	7.7%	0.00	1.50	8.5%	0.00	0.56	3.2%	0.09
Alb	mg/dl	43.87	4.0%	0.05	48.63	4.4%	0.02	0.00	0.0%	0.69
TTR	mg/dl	0.80	4.0%	0.07	3.91	19.7%	0.00	2.52	12.7%	0.00

^aSee legend, Table 1.**TABLE 3. Multiple Regression Analysis of Serum IgM/IgG Levels and Characteristics of Reference Individuals^a**

1. Step 1.

Dependent variable: IgM

Explanatory variables	Coeff. of regression	Standard error	t-value	d.f.	P
0 const	115.360	18.9780			
1 AGE	-0.8849	0.2731	3.241	486	0.00*
2 SEX	32.0888	5.3371	6.012	486	0.00**
3 SMK	-6.6031	5.3350	1.238	486	0.22

Multiple correlation coefficient R=0.355.

Step 2.

Dependent variable: IgM

Explanatory variables	Coeff. of regression	Standard error	t-value	d.f.	P
0 const	-2.0923	41.1437			
1 AGE	1.5896	0.8171	1.945	485	0.05
2 SEX	118.085	27.3127	4.323	485	0.00**
3 SMK	-7.7508	5.2967	1.463	485	0.14
4 AGE*SEX	-1.7806	0.5548	3.209	485	0.00*

Multiple correlation coefficient R=0.380.

2. Dependent variable: IgG

Explanatory variables	Coeff. of regression	Standard error	t-value	d.f.	P
0 const	1186.26	81.3535			
1 AGE	2.6666	1.1705	2.278	486	0.02
2 SEX	56.9888	22.8788	2.491	486	0.01
3 SMK	-95.203	22.8695	4.163	486	0.00**

Multiple correlation coefficient R=0.319.

^aThe results of multiple regression showing the regression coefficients for each explanatory variable, its standard error, and the statistical significance by t-test in terms of probability *P* of t-value under df or degree of freedoms. STEPs 1 and 2 in Table 3.1 correspond to the results without and with the interaction term, respectively. In consideration of the sample size and the multiple comparison nature of the analysis, the level of significance was arbitrarily set as *P* < 0.01.

P* < 0.01.*P* < 0.001.

Evaluation of Partitioning Criterion by Multiple Logistic Regression Analysis

Predominance of the relationship of common screening tests and the 13 plasma proteins with sex (SEX) or smoking habit (SMK) was analyzed using the multiple logistic regression analysis. Either SEX (Table 4.1) or SMK (Table 4.2) was

TABLE 4. Multiple Logistic Regression Analysis for the Relationship of SEX or SMK With Serum Concentrations of Major Biochemical Tests^a

1. SEX as a dependent variable

	β	SE (β)	z	P
0) const	18.7628	4.34228	4.321	(0.00)
1) UA	-1.5023	0.28013	-5.363	(0.00)**
2) GGT	-0.0495	0.01957	-2.529	(0.01)
3) RBC	-0.0386	8.614E-3	-4.476	(0.00)**
4) TCH	0.02930	9.418E-3	3.111	(0.00)*
5) IgM	0.01758	4.807E-3	3.657	(0.00)**
6) Tf	0.03621	8.757E-3	4.135	(0.00)**
7) α1AG	-0.0360	0.01718	-2.094	(0.04)
8) TTR	-0.2537	0.06064	-4.184	(0.00)**

2. SMK as a dependent variable

	β	SE (β)	z	P
0) const	-0.4401	1.15975	-0.380	(0.70)
1) SEX	-2.5151	0.35080	-7.170	(0.00)**
2) WBC	4.505E-4	1.346E-4	3.346	(0.00)**
3) IgG	-1.92E-3	7.391E-4	-2.596	(0.01)*
4) Hp	0.01071	4.102E-3	2.611	(0.01)*

^aIndependency of the relationship of major screening tests and plasma proteins with sex (SEX) or smoking habit (SMK) was analyzed. Either factor was treated as a binomial dependent variable and results of 25 common screening tests and 13 plasma proteins were set as independent variables. The final tables were shown after backward eliminations of the independent variables. The level of significance was arbitrarily set as *P* < 0.01 in consideration of the sample size and the multiple comparison nature of the analysis.

Effective number of data = 286.

P* < 0.01.*P* < 0.001.

treated as a binomial dependent variable, and results of 25 common screening tests and 13 plasma proteins were set as independent variables.

As shown in Table 4, serum IgM, Tf, and TTR were proved to have predominant relationship with SEX, whereas α_1 AG and α_2 M were not so. Regarding smoking habit, serum IgG and Hp were found to have a close association with it, together with SEX and the value of WBC, which is well known to be elevated in smokers.

Overall Decision for Partitioning Reference Individuals

Listed in Table 5 are the differences between the subgroups as a ratio to RI (%RI) after partitioning of the reference individuals by age, gender, and smoking habit. Their significance as partitioning criteria is also shown, which is based on multiple regression analysis and multiple logistic regression. The overall validity of each criterion was judged in consideration of %RI, the results of multivariate analyses, and other factors noted in Table 5.

Validity of gender as a partitioning criterion was confirmed in IgM, α_1 AG, α_2 M, and Tf. None of the 13 plasma proteins

showed apparent between-age differences when males and females were analyzed together, but in the separate analysis, serum level of IgM and Tf in the female and that of Cp in the male were obviously age-related. The validity of smoking habit as the partitioning criterion was documented in IgG. Although serum level of Hp was found to be related to smoking habit both by multiple regression analysis and by multiple logistic regression, its between-subgroup difference was < 10% of the span of RI.

DISCUSSION

There have been only a few guiding principles for partitioning reference individuals. In the NCCLS proposed guideline (1), the necessity of the partitioning is discussed and a simple statistical testing between subgroups is said to be undesirable since the test depends on the sample size: the larger the sample size, the more sensitive the test becomes. Therefore, the proposal provided a test for difference in means between subgroups and introduced an adjustment to the final judgment taking into account the sample size by the following formula:

$$z_a = \sqrt{n_1 n_2 / 240}$$

TABLE 5. Validity as a Partitioning Criterion of Age, Sex, and Smoking on Determination of Reference Intervals for 13 Plasma Proteins^a

	Age				Sex				Smoking			
	%RI	MR	MLR	VAL	%RI	MR	MLR	VAL	%RI	MR	MLR	VAL
IgG	3.2				8.7				10.1>	**	**	○
IgA	3.3				3.8				0.0			
IgM	1.2	*	—	△ ¹	12.0>	*	**	○	7.4			
C3	2.1				0.0				0.0			
C4	3.9				3.2				2.3			
Tf	4.0	*	—	△ ¹	9.4	**	**	△ ³	5.6			
α1AT	2.8				0.0				2.4			
α1AG	0.0				11.5>	**		○	8.0			
Hp	3.1				0.0				6.2	**	**	△ ³
α2M	5.8				17.7>	**		○	9.6			
Cp	7.7	*	—	△ ²	8.5				3.2			
Alb	4.0				4.4				0.0			
TTR	4.0				19.7>	**	**	○	12.7>			

^aFor 13 plasma proteins, the magnitudes of between-subgroup differences observed by partitioning criteria of age, gender and smoking habit were expressed as their percentage to the span of reference interval (%RI) in the first column. The value >10% was marked by '>'. Validity of the partitioning criterion was evaluated by multiple regression (MR) and multiple logistic regression (MLR). The latter was applied only to the validation of binomial criterion. The level of significance was expressed as '*' for $P < 0.05$ and '**' for $P < 0.01$ (the second and third columns). The overall judgment of validity (VAL) was shown in the fourth column either as △ (indeterminate) or as ○ (valid).

%RI: between-subgroup difference as a ratio to reference interval (RI) '>' signifies %RI > 10%.

MR: test of validity by the multiple regression.

MLR: test of validity by the multiple logistic regression.

*: $P < 0.01$.

***: $P < 0.001$.

—: not tested because the target variable is not binary.

VAL: overall judgment of validity as a partitioning criterion.

○: valid partitioning criterion.

△: valid partitioning criterion under some condition.

¹: valid in female.

²: valid in male.

³: valid but %RI < 10%.

However, no procedure is shown regarding a sample-size adjusted testing of differences among multiple subgroups.

In this report we adopted a three-step approach to determine whether it is appropriate to separate reference individuals according to a given criterion. The first step was to test for the statistical significance of between-subgroup difference by use of oneway ANOVA [*validation based on statistical significance*]. Because of the relatively large sample size of our study, however, even small between-subgroup differences were statistically significant. Thus the step was regarded only as an initial test of validation.

The second step was to standardize the magnitude of the difference between subgroups (after subtracting variations within the subgroups) by dividing it with the reference interval derived without subgrouping (%RI). Diagnostic (practical) significance of the difference was assessed by comparing it with %RI values for commonly measured analytes in which reference intervals are conventionally set separately for male and female. We arbitrarily adopted a criterion of %RI > 10% for partitioning reference individuals [*validation based on practical significance*]. All the between-subgroup differences that met this criterion were apparent on inspection of the distribution and were looked upon as clinically significant.

Our third step of validation was confirmation of significance of differences between subgroups by multivariate analysis [*validation by multivariate analysis*]. That is to evaluate whether a difference between subgroups specified by a partitioning criterion is truly significant in the presence of potential confounding or interacting variables. Since multiple stratification of reference individuals usually ends up with small sample sizes for subgroups, not applicable to statistical testings, the use of multiple regression analysis is a *must* to make a meaningful statistical adjustment. The formulae for the adjustment of confounder and interaction was adopted from the text of D.K. Kleinbaum (4). The result of the multiple regression analysis (Figs. 1 and 2) agreed very well with the result of stratified sex-adjusted analysis for association of smoking with serum IgG/IgM or for age-related change in IgM concentration.

Furthermore, to perform a multivariate test of validity in the reverse way, we also applied the multiple logistic regression. We set a binomial partition criterion such as gender as a

dependent variable and evaluated tests items that were expected to be associated with the criterion. The test is only supplemental and is not regarded as an essential test of validity, since the test judgment is influenced by the presence of a variable that shows a greater magnitude of between-subgroup difference. For example, a test item that has a between-sex difference of > 10% of RI can be judged insignificant by the logistic regression analysis when it is tested together with an item having a prominent between-sex difference.

The results of this study point to the necessity for additional descriptions on the validity of the partitioning criterion in the NCCLS proposed guideline. It seems essential to study the presence of confounding and interacting relationship of the partitioning criterion with other variables. Although the reference individuals may be stratified in multiple ways to clarify such relationships, the multivariate analyses are more powerful and require less sample size. Thus they are the method of choice. We also believe that the absolute magnitude of between subgroup difference is more important than the statistical judgment no matter how well the influence of sample size on the judgment is adjusted. We adopted the ratio of between subgroup difference to the span of the reference range (% RI) as a practical measure of significance for the partitioning criterion.

REFERENCES

1. NCCLS: How to define, determine, and utilize reference intervals in the clinical laboratory. *NCCLS* 12:1-47, 1992 (NCCLS:C28-P, 1992).
2. Ichihara K, Kawai T: Determination of reference intervals for 13 plasma proteins based on IFCC international reference preparation (CRM470) and NCCLS proposed guideline (C28-P, 1992): Trial to select reference individuals by results of screening tests and application of maximal likelihood method. *J Clin Lab Anal* 10:110-117, 1996.
3. Kleinbaum DG, Kupper LL, Muller KE: Confounding and interaction in regression. In *Applied Regression Analysis and Other Multivariable Methods*, 2nd ed. Duxbury Press, CA, 1988, p. 163-180.
4. Kleinbaum DG: Modeling strategy for assessing interaction and confounding. In *Logistic Regression*. Springer-Verlag, New York, 1994, p. 191-266.
5. Fisher LD, Van Belle G: In *Biostatistics: A Methodology for the Health Sciences*. J Wiley & Sons, New York, 1993, p. 496-595.
6. Baudner S: Why a new international reference preparation (IRP) for human plasma proteins? *J Clin Lab Anal* 7:273-277, 1993.
7. Sokal RR, Rohlf FJ: *Biometry*, 3rd ed., W.H. Freeman, New York, 1981, p. 207-217.