J Clin Epidemiol Vol. 49, No. 8, pp. 907–916, 1996

ELSEVIER

0895-4356/96/$15.00
PII S0895-4356(96)00025-X

Inappropriate Use of Bivariable Analysis to Screen Risk Factors for Use in Multivariable Analysis

Guo-Wen Sun, Thomas L. Shook, and Gregory L. Kay*
THE HEART INSTITUTE, GOOD SAMARITAN HOSPITAL, LOS ANGELES, CALIFORNIA 90017-2395, UNITED STATES

ABSTRACT. The use of bivariable selection (BVS) for selecting variables to be used in multivariable analysis is inappropriate despite its common usage in medical sciences. In BVS, if the statistical *p* value of a risk factor in bivariable analysis is greater than an arbitrary value (often $p = 0.05$), then this factor will not be allowed to compete for inclusion in multivariable analysis. This type of variable selection is inappropriate because the BVS method wrongly rejects potentially important variables when the relationship between an outcome and a risk factor is confounded by any confounder and when this confounder is not properly controlled. This article uses both hypothetical and actual data to show how a nonsignificant risk factor in bivariable analysis may actually be a significant risk factor in multivariable analysis if confounding is properly controlled. Furthermore, problems resulting from the automated forward and stepwise modeling with or without the presence of confounding are also addressed. To avoid these improper procedures and deficiencies, alternatives in performing multivariable analysis, including advantages and disadvantages of the BVS method and automated stepwise modeling, are reviewed and discussed. J CLIN EPIDEMIOL 49;8:907–916, 1996.

KEY WORDS. Risk factors, bivariable analysis, confounding, variable selection, bias, multivariable analysis

INTRODUCTION

In clinical research, many efforts have been made to use multivariable analysis in assessing the importance of patient characteristics or risk factors (independent variables) on outcomes (dependent variables) such as mortality, morbidity, and costs [1–61]. To identify the most significant risk factors associated with an outcome, selection of variables must occur. There are well-established statistical techniques and procedures for the automated selection of variables in multiple regression analysis, such as the backward elimination technique [62–67]. One advantage in multivariable model fittings is the reduction or elimination of confounding.

To investigate how individual risk factors may influence outcomes, bivariable analysis may be performed. A bivariable analysis defines the relationship between one independent variable and one dependent variable. In the current health care literature (1989–1994), many investigators report using the bivariable selection (BVS) method initially to select independent variables to be used in multivariable analysis [1–58]. That is, if the statistical *p* value of a risk factor is greater than a desirable arbitrary value selected by the investigators in the bivariable analysis, then this factor will not be allowed to compete for inclusion in the multivariable analysis. However, because the BVS method cannot properly control for possible confounding, using this method to screen variables for subsequent multivariable analysis introduces a source of significant error and may cause the rejection or inclusion of inappropriate variables in the multivariable analysis.

Confounding refers to the mixing of effects of risk factors on outcome and is a fundamental problem of causal inference. It biases the estimation of the effect of a risk factor on outcomes. Confounding may occur in any type of study, especially cross-sectional studies, although it is usually not a major source of bias in well-conducted prospective double-blinded randomized experiments. When a confounder exists and is not properly controlled, the estimate of the effect of a risk factor on outcome is misrepresented, that is, bias occurs. To reduce such bias, well-established statistical techniques or procedures such as the automated backward elimination procedure or stepwise regression should be used for variable selection. Because each bivariable analysis involves only one risk factor and ignores all other risk factors, it is impossible to control for confounding using bivariable analysis. Therefore, if confounding exists and the BVS method is used, estimation of effects will be biased and the corresponding conclusions distorted.

Although it seems logical to infer that a nonsignificant risk factor in bivariable analysis is likely to remain nonsignificant in multivariable analysis, there are exceptions. If risk factors ("independent variables") are not truly independent of each other, a nonsignificant risk factor in bivariable analysis is not necessarily nonsignificant in multivariable analysis. References in the literature show real world examples in which bivariable analysis failed to identify variables that are significant in multivariable analysis [59,60]. If any confounding exists, data analysis must be properly controlled for the confounder or confounders [68].

In this article, we present both hypothetical and actual data to show that the BVS method of selecting variables for multivariable analysis is inappropriate and may produce inaccurate determinations of the contributions of independent variables to outcome. In addition, alternatives including advantages and disadvantages to the BVS method are cautiously reviewed and discussed in performing multivariable analysis. In this project, all statistical analyses were performed using the SAS system, version 6.08 for Windows [69].

*Address reprint requests to Dr Sun's current address: Guo-Wen Sun, 718 Willowbrook Drive, Norristown, PA 19403.

The results, analyses, and opinions reported herein are those of the authors and in no way reflect on the opinions or judgments of the hospital, the B.C. McCabe Foundation, or The Los Angeles Thoracic and Cardiovascular Foundation.

Accepted for publication on 13 November 1995.

TABLE 1A. Hypothetical data and results (n = 4207)

Physical activity									
Inactive				Active			Inactive + active		
Patient sex		CAD	Non-CAD		CAD	Non-CAD		CAD	Non-CAD
	Female	80	800	Female	33	1200	Female	113	2000
	Male	84	1200	Male	10	800	Male	94	2000
	Odds ratio = 1.429 p = 0.028			Odds ratio = 2.200 p = 0.026			Odds ratio = 1.202 p = 0.198		

TABLE 1B. Multivariable logistic regression analysis for hypothetical data (n = 4207)

Risk factor	Odds ratio	p^a
Patient sex: Female vs. male	1.543	0.0030
Physical activity: Inactive vs. active	4.160	0.0001

[a]p value is derived by multivariable stepwise logistic regression (p = 0.10 to enter variables into the model and p = 0.05 to remove variables from the model).

TABLE 1C. Bivariable logistic regression analysis for hypothetical data (n = 4207)

Risk factor	Odds ratio	p^a
Patient sex: Female vs. male	1.202	0.1984
Physical activity: Inactive vs. active	3.814	0.0001

[a]p value is derived by bivariable logistic regression.

EXAMPLES

Hypothetical Examples

EXAMPLE 1. A cross-sectional study is conducted to estimate the effect of patient sex on the occurrence of coronary artery disease (CAD) in middle-aged women and men. Physical activity data on 4207 patients are analyzed. Because physical inactivity is a known risk factor for the CAD, we control for physical activity as a confounder, using stratified analysis. Stratified and unstratified odds ratios (ORs) are summarized in Table 1A. According to Table 1A, the unstratified analysis shows that there is no significant association between patient sex and CAD (OR = 1.202, p = 0.198). However, the stratum-specific estimates of odds ratios (OR = 1.429, p = 0.028 and OR = 2.200, p = 0.026) suggest that significant differences in incidence of CAD exist between women and men. Because the unstratified odds ratio (1.202) is less than the stratum-specific estimates (1.429 and 2.200) and different from the standardized estimate (sOR = 1.552, taking the inactive group as the standard population), we can conclude that physical activity negatively confounds the observed effect of patient sex on CAD. In fact, female sex and physical inactivity are inversely correlated (OR = 0.450, p < 0.001). Thus, the unstratified estimate is biased, and we would infer that female patients are about 43 or 120% more likely to develop CAD than male patients among physically inactive patients or among physically active patients, respectively. To confirm the existence of the confounder, an alternative method, logistic regression, is applied to the same data, and we see in Table 1B the same conclusion: sex and physical activity independently affect the occurrence of CAD.

However, the bivariable analysis (Table 1C) suggests that the effect of patient sex on CAD is not statistically significant (p = 0.198), which is the same conclusion reached in the unstratified analysis. In applying the BVS method to this data set, we miss the chance to explore the true relationship between patient sex and CAD, which is demonstrated in Table 1A by stratified analysis and in Table 1B by stepwise logistic regression. The true relationship can be detected only by properly controlling the confounder, physical activity, using either stratified analysis or multivariable logistic regression of CAD on both factors of physical activity and patient sex.

To further illustrate this phenomenon, we use a hypothetical model to describe the effect of confounding on data analysis. Assume that the true risk difference in CAD between women and men consists of two major components: observed and confounded. Because female sex is negatively confounded by physical inactivity, the observed difference is then reduced by the confounder, physical activity, and is biased away from the true difference. If we fail to control for physical activity as a confounder, the observed difference between the two sexes is not enough to show a significant difference in bivariable analysis. The only way to obtain a less biased estimate of the true difference is to control for the confounder.

It has been argued that perhaps the use of a larger p value, say p = 0.15, through the BVS method can solve this problem. In the literature, some investigators have used the p = 0.10 level, others have used the p = 0.20 level [1–57]. In the book by Hosmer and Lemeshow, the level of p = 0.25, which is the largest value among articles we reviewed, is recommended as a screening criterion for selection of candidate variables to be used in multivariable analysis through the BVS method [58]. As shown in Table 2A and B, use of the BVS method leads one to conclude that patient sex is not significantly associated with CAD (OR = 1.176, p = 0.267) and, even with a significance level as high as 0.25, sex will not be included as a candidate variable in the multivariable analysis. However, as demonstrated in Table 2A by stratified analysis and Table 2C by multivariable analysis, sex is found to be significantly associated with CAD (OR = 1.503, p = 0.0064), after controlling for the confounding.

EXAMPLE 2. In Table 3A we slightly modify the numbers shown in Table 1A. The unstratified analysis shows that there is no significant association between patient sex and CAD (OR = 1.25, p = 0.112).

TABLE 2A. Hypothetical data and results (n = 4068)

Physical activity									
Inactive				**Active**				**Inactive + active**			
Patient sex		CAD	Non-CAD		CAD	Non-CAD		CAD	Non-CAD
	Female	75	774	Female	32	1161	Female	107	1935
	Male	81	1161	Male	10	774	Male	91	1935
	Odds ratio = 1.389 p = 0.048			Odds ratio = 2.133 p = 0.034			Odds ratio = 1.176 p = 0.267		

TABLE 2B. Bivariable logistic regression analysis for hypothetical data (n = 4068)

Risk factor	Odds ratio	p^a
Patient sex: Female vs. male	1.176	0.2679
Physical activity: Inactive vs. active	3.714	0.0001

[a] p value is derived by bivariable logistic regression.

TABLE 2C. Multivariable logistic regression analysis for hypothetical data (n = 4068)

Risk factor	Odds ratio	p^a
Patient sex: Female vs. male	1.503	0.0064
Physical activity: Inactive vs. active	4.030	0.0001

[a] p value is derived by multivariable stepwise logistic regression (p = 0.10 to enter variables into the model and p = 0.05 to remove variables from the model).

However, the stratum-specific estimates of odds ratios (OR = 1.756, p = 0.015 and OR = 1.095, p = 0.634) suggest that significant differences in incidence of CAD exist between women and men in the physically active group. Because the unstratified estimate of odds ratio (OR = 1.25) differs from both stratum-specific estimates (1.095 and 1.756) and the latter two estimates differ from each other, we assess possible confounding by comparing the unstratified estimate to a standardized estimate. In this example, the standardized estimate of the odds ratio (sOR = 1.383) also differs from the unstratified as well as the stratum-specific estimates. From this we can infer that physical activity is a confounder. In addition, Table 3A possesses a very important characteristic: symmetry with respect to factors of patient sex and physical activity. This symmetry implies that discussions for physical activity would also apply to patient sex.

TABLE 3A. Hypothetical data and results (n = 4016)

Physical activity									
Inactive				**Active**				**Inactive + active**			
Patient sex		CAD	Non-CAD		CAD	Non-CAD		CAD	Non-CAD
	Female	50	750	Female	70	1150	Female	120	1900
	Male	70	1150	Male	26	750	Male	96	1900
	Odds ratio = 1.095 p = 0.634			Odds ratio = 1.756 p = 0.015			Odds ratio = 1.250 p = 0.112		

TABLE 3B. Bivariable logistic regression analysis for hypothetical data (n = 4016)

Risk factor	Odds ratio	p^a
Patient sex: Female vs. male	1.250	0.1128
Physical activity: Inactive vs. active	1.250	0.1128

[a] p value is derived by bivariable logistic regression.

TABLE 3C. Multivariable logistic regression analysis for hypothetical data (n = 4016)

Risk factor	Odds ratio	p^a
Patient sex: Female vs. male	1.329	0.0486
Physical activity: Inactive vs. active	1.329	0.0486

[a] p value is derived by multivariable backward logistic regression (p = 0.05 to remove variables from the model).

TABLE 4A. Hypothetical data and results (n = 600)

Physical activity								
	Inactive			Active			Inactive + active	
Patient sex		Diseased	Non-D[a]		Diseased	Non-D[a]		Diseased	Non-D[a]
	Female	80	74	Female	32	120	Female	112	194
	Male	84	120	Male	10	80	Male	94	200
	Odds ratio = 1.544 p = 0.043			Odds ratio = 2.133 p = 0.048			Odds ratio = 1.228 p = 0.233		

[a]Non-D, nondiseased.

TABLE 4B. Bivariable logistic regression analysis for hypothetical data (n = 600)

Risk factor	Odds ratio	p^a
Patient sex: Female vs. male	1.228	0.2329
Physical activity: Inactive vs. active	4.026	0.0001

[a]p value is derived by bivariable logistic regression.

TABLE 4C. Multivariable logistic regression analysis for hypothetical data (n = 600)

Risk factor	Odds ratio	p^a
Patient sex: Female vs. male	1.670	0.0060
Physical activity: Inactive vs. active	4.534	0.0001

[a]p value is derived by multivariable stepwise logistic regression (p = 0.10 to enter variables into the model and p = 0.05 to remove variables from the model).

That is, significant differences in the incidence of CAD exist between the two physical activity groups for males (OR = 1.756, p = 0.015), but not for females (OR = 1.095, p = 0.634).

We again invite readers to pay special attention to the fact that neither physical activity nor patient sex is significantly correlated with CAD by using the BVS method (Table 3A and B) for analysis of these data. With appropriate statistical technique, however, we found that both factors are mutually confounders. In fact, female sex and physical inactivity are inversely correlated (OR = 0.417, $p < 0.001$). If we had used the BVS method prior to performing multivariable analysis, we would have had no chance to explore the true relationship between patient sex, physical activity, and CAD.

Again multivariable logistic regression, like stratified analysis, confirmed the mutual confounding of sex and physical activity. Results are shown in Table 3A and C, respectively. Here, we should emphasize that Table 3C cannot be obtained by using either forward or stepwise variable selection techniques with any setting of p value < 0.11275 to enter variables into the model or any setting of p value < 0.11275 to remove variables from the model. For example, if one applies stepwise logistic regression to the data shown in Table 3A and set p value = 0.15 to enter variables into the model and p value = 0.10 to remove variables from the model, he or she will never have a chance to see Table 3C. However, Table 3C can be obtained by applying (1) the forward selection method with any setting of p value > 0.11285 to enter variables into the model, or (2) the stepwise selection method with any setting of p value > 0.11285 to both enter variables into and remove variables from the model, or (3) the backward elimination method with p value = 0.05 to remove variables from the model. This example demonstrates one of the advantages of using the backward elimination procedure, compared to both the forward and stepwise selection methods, in multiple regression analysis. A more theoretical but less realistic example can be found in the paper by Mantel [63].

In the construction of the hypothetical examples shown above (Tables 1 to 3), the sample sizes used were moderately large. One of the main arguments is to minimize the gap between hypothetical examples and actual observations. The incidence of a disease such as CAD used in our hypothetical examples is relatively low (<5%), thus a moderately large sample size is necessary to make any appropriate and meaningful inferences statistically. If we use a different outcome with high incidence rate, another hypothetical example similar to Example 1 above but with much smaller sample size can be constructed as shown in Table 4 (n = 600, proportion of diseased patients = 34.3%). However, interpretation of Table 4 is limited and cannot reasonably be linked to incidence of CAD in a general population.

Actual Examples

To show how the BVS method may generate wrong results in practice, examples are drawn from the multivariable prediction of hospital mortality for cardiac patients.

EXAMPLE 1. To study the influence of prior thrombolytic therapy on hospital mortality, data were reviewed on 3359 consecutive cardiac patients referred to a nonprofit hospital for diagnosis and/or interventional procedures (including cardiovascular surgery). Information such as age and sex, transport mode, in-transport treatments and conditions, preadmission diagnosis, and history of prior thrombolytic therapy (Table 5) was collected. Overall, 379 (11.3%) patients received thrombolytic therapy prior to transport. To determine how the above factors affected hospital mortality, both bivariable and stepwise multivariable logistic regression were performed on the same data. As displayed in Table 5, the p value for the effect of prior thrombolytic therapy on hospital mortality was 0.1698 in the bivariable analysis but became 0.0013 in the multivariable analysis. If, initially, we had applied the BVS method at a significant p value of 0.05, 0.10, or even 0.15, prior thrombolytic

TABLE 5. Risk analysis of hospital mortality: Referral database $n = 3359$

Risk factors	Bivariable		Multivariable	
	OR[a]	p[a]	OR[a]	p[a]
Age (in 10-year interval)	1.682	0.0001	1.726	0.0001
Sex (female vs. male)	1.525	0.0058	1.337	0.0931
Transport mode (helicopter vs. others)	5.502	0.0001	2.474	0.0010
Diagnosis of:				
Aortic dissection/aneurysm	6.852	0.0001	7.696	0.0001
Angina (stable/unstable)	0.357	0.0001	NS	
Atrial tachyarrhythmia	NS		NS	
Heart failure/valvular heart disease	2.218	0.0002	NS	
Ventricular tachycardia/fibrillation	NS		NS	
Myocardial infarction	1.686	0.0007	1.948	0.0004
During transport:				
IABP support	10.10	0.0001	NS	
Pressors	9.702	0.0001	2.701	0.0001
i.v. nitrates	NS		NS	
i.v. antiarrhythmics	2.299	0.0001	1.460	0.0838
Heparin	NS		NS	
Ventilator/CPR/defibrillator	12.64	0.0001	2.456	0.0037
Thrombolytics given pre-transport	0.685	0.1698	0.347	0.0013
Pretransport ischemic discomfort	NS		NS	
Heart failure (rales/S3)	4.310	0.0001	2.397	0.0001
During transport:				
Minimum blood pressure not in 110–149 mmHg	2.819	0.0001	2.133	0.0001
Minimum heart rates > 90 beats/minute	3.105	0.0001	NS	
Maximum heart rates > 100 beats/minute	3.522	0.0001	1.850	0.0016

[a]By bivariable or stepwise logistic regression ($p = 0.15$ to enter variables into the model and $p = 0.10$ to remove variables from the model). OR, odds ratio; NS, not significant ($p > 0.15$).

therapy would not have been allowed to compete for inclusion in the multivariable analysis.

To investigate whether the factor prior thrombolytic therapy was confounded by some other risk factor(s), we examined its association with all other risk factors included in this study. Other factors that are positively and statistically associated ($p < 0.05$) with prior thrombolytic therapy include treatments during transport (pressors, intravenous [i.v.] antiarrhymics, intra-aortic balloon pump [IABP], heparin, i.v. nitrates), diagnosis of myocardial infarction (MI), transport via helicopter, minimum systolic blood pressure not in range of 110–149 (mmHg), pretransport ischemic discomfort, and heart rate (beats per minute). In addition, prior thrombolytic therapy is negatively ($p < 0.05$) associated with the risk factors of aortic dissection/aneurysm, diagnoses (angina, ventricular tachycardia/fibrillation, heart failure/valvular heart disease), female sex, and advancing age.

Up to now it has been difficult to judge how prior thrombolytic therapy is confounded by the joint effects of these several positively and negatively correlated factors. If we can show that patients who receive prior thrombolytic therapy are at high risk of hospital mortality, then we can conclude that prior thrombolytic therapy is positively confounded. To explore this concept, a risk index score is calculated, patient by patient, on the basis of the estimated odds ratios in multivariable analysis. Only risk factors with a p value $<$ 0.10 (Table 5) are included in the calculation. For a patient, the risk index score is defined as the total sum of the logarithm of the estimated (multivariable) odds ratios (except for the therapy) reported in Table 5. For example, a patient who was 65 years old and used pressors during transport will have a risk index score of 4.56 [i.e., 65 × log(1.73)/10(age 65) + log(2.70) (pressors)]. Thus, the average risk index score is significantly higher in patients treated with prior thrombolytic therapy (4.97, median = 4.75) than in patients untreated with such therapy (4.44, median = 4.33, $p < 0.001$ by t test, Wilcoxon test, and Kolmogorov–Smirnov test).

Thus, we can conclude that the prior thrombolytic therapy is confounded with other risk factors in the data set.

EXAMPLE 2. To determine the influence of preoperative risk factors on hospital mortality, data were reviewed from 1355 consecutive isolated coronary artery bypass graft (CABG) patients during a 28-month period. The following information was collected preoperatively for each patient: age, sex, body surface area, weight index (ratio of actual weight to ideal weight), ejection fraction, onset of myocardial infarction, emergent status of procedure, prior CABG, history of diabetes, and preoperative use of intra-aortic balloon pump (IABP). The type of cardioplegia (warm vs. cold blood), the type of graft (vein alone vs. any internal mammary artery), and the number of grafts were also used to adjust estimates of parameters. Comparison of both biovariable and multivariable analysis of hospital mortality is presented in Table 6. As shown in Table 6, the p value for the prior CABG is 0.1334 by bivariable analysis but 0.0253 by multivariable analysis. Furthermore, the statistical p value for the history of diabetes was $p = 0.2426$ (definitely not significant) by bivariable analysis but 0.1313 (marginally significant) by multivariable analysis.

To show that estimates derived from the bivariable analyses are inaccurate due to confounding, we examined the association between prior CABG and history of diabetes and the other risk factors listed in Table 6. The history of CABG is significantly correlated with large body surface area (BSA) ($p = 0.012$), middle age (55–

TABLE 6. Risk analysis of hospital mortality: CABG Database $n = 1355$

Risk factors	Bivariable		Multivariable	
	OR[a]	p^a	OR[a]	p^a
Age (in 10-year interval)	1.697	0.0006	1.586	0.0079
Sex (female vs. male)	NS		NS	
Body surface area:				
Large (≥2.0)	1.000	0.0029	1.000	0.0541
Medium (1.6–1.99)	2.044		1.687	
Small (<1.6)	4.179		2.845	
Weight index:				
Normal (1.0–1.59)	1.000	0.0081	1.000	0.0124
Cachexia (<1.0)	1.701		1.748	
Obesity (≥1.6)	2.893		3.057	
Emergency procedure	3.008	0.0001	1.718	0.0888
Preoperative IABP support	8.073	0.0001	3.120	0.0070
Previous CABG	1.580	0.1334	2.088	0.0253
History of diabetes	1.437	0.2426	1.665	0.1313
Preoperative ejection fraction < 40%	4.179	0.0001	2.423	0.0039
Prior MI days:				
None	1.000	0.0001	1.000	0.0004
≥15 days	1.698		1.440	
8–14 days	2.884		2.074	
3–7 days	4.898		2.987	
1–2 days	8.318		4.303	
Vein graft alone	4.175	0.0005	NS	
Number of grafts	NS		NS	
Cardioplegia technique (warm vs. cold blood)	NS		NS	

[a]By bivariable or stepwise logistic regression ($p = 0.15$ to enter variables into the model and $p = 0.10$ to remove variables from the model). The factor of history of diabetes was forced into the final model. OR, odds ratio; NS, not significant ($p > 0.15$).

75 years) ($p < 0.001$), male sex ($p = 0.005$), and lesser number of grafts ($p < 0.001$), which (except for age) are generally believed to be lesser in risk than small BSA, female sex, and large number of grafts, respectively. However, a procedure involving vein graft alone is more likely to be performed for patients with a history of CABG than for patients without such a history ($p < 0.001$). Note that the type of graft used is not randomly assigned to patients and vein graft alone is commonly used for patients at higher risk. Because patient sex, type of graft used, and number of grafts are not statistically significant ($p > 0.15$) in the multivariable analysis and BSA and age are statistically associated with hospital mortality ($p = 0.0541$ for BSA and $p = 0.0079$ for age), it can be inferred that the history of CABG is negatively confounded by the factors of BSA and patient age. Logically, the factors BSA and age are confounders. To reduce bias and to eliminate confounding, the factors of prior CABG and the two confounders (BSA and age) should be simultaneously included in the multivariable analysis.

To investigate whether the history of diabetes is also confounded with other risk factors, we examined distributions of risk factors in diabetic and nondiabetic patients. Comparing distribution of patients within the BSA groups, only 55.6% of the diabetic patients had medium BSA (1.6–1.99), but 61.8% of nondiabetic patients had medium BSA ($p = 0.058$). The same pattern is observed for patient weight index and history of MI. For weight index, only 15.9% of diabetic patients but 25.2% of nondiabetic patients are cachexic (weight ratio < 1.0, $p = 0.001$). With the diagnosis of MI, there are only 15.1% of diabetic patients but 20.1% of nondiabetic patients ($p = 0.060$). On the other hand, history of diabetes is positively correlated with low ejection fraction ($p = 0.001$), female sex ($p < 0.001$), and large number of grafts ($p = 0.043$). To determine how these risk factors confounded the history of diabetes, the risk index score is also calculated for each patient. There is no significant difference between diabetic patients and nondiabetic patients for the multivariable risk index scores (4.22 vs. 4.27, $p = 0.446$). On the basis of the above information, we conclude that the confounding between the history of diabetes and other risk factors is negative but not strong. This conclusion is supported by the small decrease in the p value shown in Table 6.

COMMENTS AND CONCLUSIONS

We believe that the use of the BVS method in multivariable analysis is one of the most common errors in data analysis in the current literature. Many investigators now perform multivariable analysis based on variables selected only through the BVS method [1–58]. In this article both hypothetical and actual examples have been presented to show that the use of the BVS method to select candidate variables will produce incorrect estimates if confounding problems are not properly controlled. With bivariable screening, variables that failed the initial screening are ignored when confidence intervals, standard errors, p values, and R^2 are computed. Altman and Andersen investigated the stability of a Cox regression model using the bootstrap technique and concluded that the confidence intervals of the estimated survival probabilities for the individual subjects based on the stepwise regression models were considerably wider than those based on the models with the variables fixed in advance [70]. In fact, with the BVS method, multivariable regression models are fitted at a preselected subset of candidate variables, and therefore are more likely to lead to overopti-

mistic fitting if confounding is not properly controlled [70]. If selection of variables must be performed, standard statistical procedures but not the BVS method should be used.

It has been argued that there are several disadvantages and/or problems in selection of variables using standard procedures in multivariable analysis, especially in using automated forward and stepwise regression [62,71,72]. First, if missing values exist in some of the variables or fields and a large number of these incomplete variables are to be included in the calculations without imputation, the effective sample size will be significantly reduced in multivariable estimations and may cause distorted analysis. Although missing data can theoretically be filled by applying modern imputation techniques, most of these methods are either too simple (such as mean substitution) or too complicated (such as maximum likelihood). Even if the regression equation method seems to be more practically suitable and acceptable for nonstatisticians, it has not yet been popularly used by investigators in the medical sciences [73]. In addition, the proportion of missing values associated with a specific variable or risk factor must be reasonably small in order to apply meaningful imputation. However, any application of such imputation methods may result in biases of unknown magnitude in the estimates of parameters [66]. At this point, we certainly agree with Afifi and Clark that the most understandable analysis is the one based on a complete data set only [66]. To obtain a complete data set that is as large as possible, the investigator may have to delete some variables from the analysis. In this article, we have shown that such decision making must be very cautious and should not be solely based on the BVS method.

Second, the greater the number of variables to be included in a model fitting, the more computer resources (e.g., memory or time) are required to obtain estimates of parameters by using automated variable selection procedures. This concern is becoming less problematic as computer technology and statistical software packages are upgraded [69].

Third, if supposedly "independent" variables are closely related to one another the input data yield collinearity or multicollinearity so that estimates of parameters will be unstable in the sense that small changes in the data may cause large changes in the estimates, or the estimates obtained may be nonsense due to rounding errors in computations. However, this is not usually a serious problem with modern statistical computing techniques [62]. In addition, most sophisticated statistical software packages are designed to include diagnostic features for detecting collinearity [69]. If a collinearity exists, it is a good idea to find out which variables are nearly collinear with which other variables and, in most of the cases, the investigator may have to combine the collinear variables using methods such as analysis of principal components or to decide which variables should be removed from the analysis. The latter type of variable screening should be based on prior knowledge of the variables or principles of the subject matter investigated. The BVS method in no way solves any problem caused by the collinearity because it deals with each independent variable separately.

Finally, when the pool of independent variables contains both authentic variables and noise variables, the automated variable selection in multiple regression analysis often selects a sizable percentage of noise variables [71,72]. Several research reports from simulation studies recommend setting the p value from 0.15 to 0.25 for entering variables and setting the p value from 0.10 to 0.15 for deleting variables in the automated variable selection procedures in order to maximize the proportion of the authentic variables [71,72,74–76]. An alternative way to avoid selection of noise variables is to select the initial set of independent variables carefully, including for study purposes only those variables that are known or expected to be risk factors of the outcome either by principles in the study area of the subject matter or by prior knowledge such as previous research findings [71,72]. Because the BVS method is unable to distinguish authentic variables from noise variables due to confounding, use of the BVS technique will be misleading. In this article, we have shown in both hypothetical and actual examples that authentic variables may be mistakenly eliminated by the BVS method.

ALTERNATIVES TO THE BVS METHOD. An effective and practical strategy for dealing with the confounding problems addressed in this article is to perform stratified analysis when the number of risk factors is not too large or, if possible, to fit the full model with all candidate variables, deleting only those variables whose coefficients are clearly insignificant and have a sign that makes no sense. Alternatively, if standard statistical procedures are preferred, the automated backward elimination regression can be used [62–78]. Because the forward and stepwise regression are highly sensitive to specification of their stopping rules in the presence of confounding, as demonstrated in Example 2 under Hypothetical Examples (above) and in Mantel, both the full-model fits and the backward elimination regression are two satisfactory procedures, especially for investigators who like to see all of the candidate variables in the equation once, in order "not to miss anything" [62,63]. In addition, there are other advantages of full-model fits or backward regression such as correcting standard errors, p values, confidence intervals, and proper documentation of all variables considered. Furthermore, both the full-model fits and the backward elimination procedure are much more economical of effort and efficiencies than are the other two procedures frequently used by clinicians in practice [63].

Although the stratified analysis, the full-model fit, and the backward elimination procedure have their own limitations or disadvantages, compared to others, these methods provide great opportunities to control for potential confounding. In the case that the data set does not support beginning the multivariable analysis with all possible independent variables, certain steps must be followed.

1. If collinearity exists, statistical packages such as the SAS system can be used to find out which variables are nearly collinear with which other variables [69]. If the investigator wishes to screen variables in order to avoid collinearity, this type of screening must be based on prior knowledge of the variables or principles of the subject matter studied; otherwise, special statistical analysis such as Ridge regression or analysis of principal components should be performed [62]. Here, we should point out that the investigator should not let a computer program decide what is to be done about collinearity. Additionally, the BVS method is useless to solve any problem caused by collinearity.
2. If the ratio of number of cases or patients to number of independent variables is too small in the data, the investigator should begin to select the initial set of independent variables carefully. Only those variables that are known or expected to be risk factors for the outcome by either principles of the study or prior knowledge of the variables should be included. Importantly, the BVS method should never be used to solve this problem.
3. Special attention should be paid to any independent variables resulting in a zero cell in the contingency tables with discrete outcome (dependent) variable. If the independent variable has more than two categories, the simplest way to handle this prob-

lem is to eliminate the zero cell by collapsing the categories of the variable in some meaningful ways, or to treat the variable as if it were continuous if the variable is ordinal scaled [58]. If the independent variable has only two categories, then the variable must be treated cautiously in the modeling stage of the analysis [58]. Another similar issue is that there are only a few observations in any one of the four cells such that both the estimated coefficient and its standard error are unreasonably large or unstable, so that the final model is very sensitive to the inclusion of such variables. Although these variables may be forced to be excluded from the final model, analysis or conclusions made must be incorporated appropriately.

4. If steps 1 to 3 above still cannot solve the convergency problems in estimations, attempts should be made to find out which variable or variables caused the problem. If necessary, the automated stepwise or forward selection procedures can be used with various stopping rules. Simulation studies recommend setting the p value from 0.15 to 0.25 for entering variables into the model and setting the p value from 0.10 to 0.15 for deleting variables from the model [71,72,74–76].

No matter what statistical methods are to be used in data analysis, it should always be remembered that testing associations between variables in the data cannot demonstrate that a particular variable is or is not a confounder because statistical significance does not indicate the magnitude of the association. In fact, the statistical p value is highly correlated with the sample size and a small sample size is more likely to produce large p values. In general, variables should not be selected solely on the basis of statistical data analysis and the automated variable selection methods should serve only as ancillary tools in interpretation of the data [71,72]. In the intermediate stages of analysis, the investigator can consider several equations as suggested by one or more of the variable selection procedures with various stopping rules [66]. In some cases, independent variables may be forced into the model according to their clinical importance even if they did not achieve statistical significance [79]. Selected variables in the final model, in which conclusions are to be drawn, should be critically evaluated through either confirmation in the subject matter or validation in separate data sets [66,72].

In multivariable analysis, there are two types of questions to be addressed in a specific research project: (1) the assessment of the prognostic effect of a risk factor considered, and (2) the formulation of a risk index or score aiming at an optimal discrimination between patients with low and high risk associated with an outcome. With a risk index, statistically it is true for a given population (or data base) that the choice of variables and their individual (possibly confounded) effects is of limited importance as long as the risk classification of patients is unaffected. However, there is no control over whether the risk classification of patients will be affected by applying a "confounded" index (e.g., a "right" variable is replaced by a confounder) to another population. Therefore, the assessment of the prognostic effect of each individual risk factors and the right choice of variables (corrected for confounding) are prerequisite for the formulation of the risk index in a multivariable analysis. Consequently, confounding is a matter of concern not only for the assessment of individual risk factors but also for the formulation of a risk index, given the consideration of the causal net underlying the risk factors under study.

Another issue we should mention is that assessment of variable importance should not be based solely on statistical analysis or procedures. Particularly, this assessment should not be based on the inclusion/noninclusion of variables in the final model or on the order of the entry/deletion of variables into/from the final model [71,80]. Our findings in this article strongly support this principle, directly and/or indirectly, that the last variable entered into the final model is not necessarily the least important one in the data set (see Example 2 under Hypothetical Examples).

In summary, the BVS method frequently used by many investigators is inappropriate for selecting variables to be used in multivariable analysis. Examples based on hypothetical and actual data are presented to support this statement. The primary argument for not using the BVS method is that it cannot properly control confounding or intercorrelations between independent variables. Confounding may occur in any type of study, especially in cross-sectional studies. Thus, when a confounder exists and is not properly controlled, estimation of the effects of a risk factor on outcome is biased and distorted. Because the BVS method is unable to correct for confounders, the use of the BVS method is not an appropriate way to select variables to be used in multivariable analysis. To reduce potential bias and to eliminate confounding, well-established statistical methods such as the full-model fit, the backward elimination procedure, and stepwise regression with various stopping rules are recommended for variable selection in multivariable analysis.

This project would not have reached this stage without critical review from one of the anonymous referees. The authors thank Olivia A. Crookes, M.D., for her helpful comments and suggestions. The authors also gratefully acknowledge the assistance of Lucille N. Junio, RN, MSN, and Irma Larsen for preparation of the databases used in this project. The first author (G.W.S.) is partially supported by the B.C. McCabe Foundation (Whittier, CA) and The Los Angeles Thoracic and Cardiovascular Foundation (Los Angeles, CA).

References

1. Jen J, Kim H, Piantadosi S, *et al.* Allelic loss of chromosome 18q and prognosis in colorectal cancer. **N Engl J Med** 1994; 331: 213–221.
2. Muss HB, Thor AD, Berry DA, *et al.* c-erbB-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. **N Engl J Med** 1994; 330: 1260–1266.
3. Small PM, Hopewell PC, Singh SP, *et al.* The epidemiology of tuberculosis in San Francisco—a population-based study using conventional and molecular methods. **N Engl J Med** 1994; 330: 1703–1709.
4. Fiatarone MA, O'Neill EF, Ryan ND, *et al.* Exercise training and nutritional supplementation for physical frailty in very elderly people. **N Engl J Med** 1994; 330: 1769–1775.
5. Cook DJ, Fuller HD, Guyatt GH, *et al.* Risk factors for gastrointestinal bleeding in critically ill patients. **N Engl J Med** 1994; 330: 377–381.
6. Cleeland CS, Gonin R, Hatfield AK, *et al.* Pain and its treatment in outpatients with metastatic cancer. **N Engl J Med** 1994; 330: 592–596.
7. Aprikian AG, Zhang ZF, Fair WR. Prostate adenocarcinoma in men younger than 50 years. **Cancer** 1994; 74: 1768–1777.
8. Murphy BA, Motzer RJ, Mazumdar, *et al.* Serum tumor marker decline is an early predictor of treatment outcome in germ cell tumor patients treated with cisplatin and ifosfamide salvage chemotherapy. **Cancer** 1994; 73: 2520–2526.
9. Newland RC, Dent OFD, Lyttle MNB, Chapuis PH, Bokey EL. Pathologic determinants of survival associated with colorectal cancer with lymph node metastases. **Cancer** 1994; 73: 2076–2082.
10. Yost LS, McCorkle R, Buhler-Wilkerson K, Schultz D, Lusk E. Determinants of subsequent home health care nursing service use by hospitalized patients with cancer. **Cancer** 1993; 72: 3304–3312.
11. Christensen ML, Hancock ML, Gattuso J, *et al.* Parenteral nutrition associated with increased infection rate in children with cancer. **Cancer** 1993; 72: 2732–2738.
12. Abrutyn E, Mossey J, Berlin JA, *et al.* Does asymptomatic bacteriuria predict mortality and does antimicrobial treatment reduce mortality in elderly ambulatory women? **Ann Intern Med** 1994; 120: 827–833.
13. Fishbein DB, Dawson JE, Robinson LE. Human ehrlichiosis in the United States, 1985 to 1990. **Ann Intern Med** 1994; 120: 736–743.
14. Shaw LJ, Miller DD, Romeis JC, Kargl D, Younis LT, Chaitman, BR. Gender difference in the noninvasive evaluation and management of

patients with suspected coronary artery disease. **Ann Intern Med** 1994; 120: 559–566.
15. Lynch A, Marlar R, Murphy J, *et al.* Antiphospholipid antibodies in predicting adverse pregnancy outcome. **Ann Intern Med** 1994; 120: 470–475.
16. Rigotti NA, McKool KM, Shiffman S. Predictors of smoking cessation after coronary artery bypass graft surgery. **Ann Intern Med** 1994; 120: 287–293.
17. Hodis HN, Mack WJ, Azen SP, *et al.* Triglyceride- and cholesterol-rich lipoproteins have a differential effect on mild/moderate and severe lesion progression as assessed by quantitative coronary angiography in a controlled trial of lovastatin. **Circulation** 1994; 90: 42–49.
18. Kimball TR, Daniels SR, Khoury PR, Magnotti RA, Turner AM, Dolan LM. Cardiovascular status in young patients with insulin-dependent diabetes mellitus. **Circulation** 1994; 90: 357–361.
19. Enriquez-Sarano M, Tajik AJ, Schaff HV, Orszulak TA, Bailey KR, Frye RL. Echocardiographic prediction of survival after surgical correction of organic mitral regurgitation. **Circulation** 1994; 90: 830–837.
20. Volpi A, Vita CD, Franzosi MG, *et al.* Determinants of 6-month mortality in survivors of myocardial infarction after thrombolysis—results of the GISSI-2 data base. **Circulation** 1993; 88: 416–429.
21. Elayda MAA, Hall RJ, Reul RM, *et al.* Aortic valve replacement in patients 80 years and older. **Circulation** 1993; 88(part 2): 11–16.
22. Sarris GE, Robbins RC, Miller DC, *et al.* Randomized, prospective assessment of bioprosthetic valve durability—Hancock versus Carpentier–Edwards valves. **Circulation** 1993; 88(part 2): 55–64.
23. Quigley RL, Milano CA, Smith LR, White WD, Rankin JS, Glower DD. Prognosis and management of anterolateral myocardial infarction in patients with severe left main disease and cardiogenic shock—the left main shock syndrome. **Circulation** 1993; 88(part 2): 65–70.
24. Wallis JB, Supino PG, Borer JS. Prognostic value of left ventricular ejection fraction response to exercise during long-term follow-up after coronary artery bypass graft surgery. **Circulation** 1993; 88(part 2): 99–109.
25. Addonizio LJ, Hsu DT, Douglas JF, *et al.* Decreasing incidence of coronary disease in pediatric cardiac transplant recipients using increased immunosuppression. **Circulation** 1993; 88(part 2): 224–229.
26. Rosengart TK, Krieger K, Lang SJ, *et al.* Reoperative coronary artery bypass surgery—improved preservation of myocardial function with retrograde cardioplegia. **Circulation** 1993; 88(part 2):330–335.
27. Yano K, Grove JS, Reed DM, Chun HM. Determinants of the prognosis after a first myocardial infarction in a migrant Japanese population—the Honolulu Heart Program. **Circulation** 1993; 88: 2582–2595.
28. Popma JJ, Satler LF, Pichard AD, *et al.* Vascular complications after balloon and new device angioplasty. **Circulation** 1993; 88(part 1): 1569–1578.
29. Talley JD, Weintraub WS, Roubin GS, *et al.* Failured elective percutaneous transluminal coronary angioplasty requiring coronary artery bypass surgery. **Circulation** 1990; 82: 1203–1213.
30. Khan SS, Kupfer JM, Matloff JM, Tsai TP, Nessim S. Interaction of age and preoperative risk factor in predicting operative mortality for coronary bypass surgery. **Circulation** 1992; 86(Suppl. II): 186–190.
31. Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. **Circulation** 1989; 79(Suppl. I): 3–12.
32. Basoglu M, Paker M, Ozmen E, Tasdemir O, Sahin D. Factors related to long-term traumatic stress responses in survivors of torture in Turkey. **JAMA** 1994; 272: 357–363.
33. Zolopa AR, Hahn JA, Gorter R, *et al.* HIV and tuberculosis infection in San Francisco's homeless adults—prevalence and risk factors in a representative sample. **JAMA** 1994; 272: 455–461.
34. Schwartz B, Fries S, Fitzgibbon AM, Lipman H. Pediatricians' diagnostic approach to pharyngitis and impact of CLIA 1988 on office diagnostic tests. **JAMA** 1994; 271: 234–238.
35. Bell MR, Holmes DR Jr, Berger PB, Garratt KN, Bailey KR, Gersh BJ. The changing in-hospital mortality of women undergoing percutaneous transluminal coronary angioplasty. **JAMA** 1993; 269: 2091–2095.
36. Higgins TL, Estafanous FG, Loop FD, Beck GJ, Blum JM, Paranandi L. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients. **JAMA** 1992; 267: 2344–2348.
37. Hannan EL, Arani DT, Johnson LW, Kemp HG Jr, Lukacik G. Percutaneous transluminal coronary angioplasty in New York State—risk factors and outcomes. **JAMA** 1992; 268: 3092–3097.
38. Williams SV, Nash DB, Goldfarb N. Difference in mortality from coronary artery bypass graft surgery at five teaching hospitals. **JAMA** 1991; 266: 810–815.
39. Kelman HR, Thomas C, Kennedy GJ, Cheng J. Cognitive impairment and mortality in older community residents. **Am J Public Health** 1994; 84: 1255–1260.
40. Kiel DP, Eichorn A, Intrator O, Silliman RA, Mor VM. The outcomes of patients newly admitted to nursing homes after hip fracture. **Am J Public Health** 1994; 84: 1281–1286.
41. Szabo BM, van Veldhuisen DJ, Crijns HJGM, Wiesfeld ACP, Hillege HL, Lie KI. Value of ambulatory electrocardiographic monitoring to identify increased risk of sudden death in patients with left ventricular dysfunction and heart failure. **Eur Heart J** 1994; 15: 928–933.
42. Boudrez H, Backer GD, Comhaire B. Return to work after myocardial infarction: Results of a longitudinal population based study. **Eur Heart J** 1994; 15: 32–36.
43. Graffagnino C, Gasecki AP, Doig GS, Hachinski VC. The importance of family history in cerebrovascular disease. **Stroke** 1994; 25: 1599–1604.
44. Anderson CS, Jamrozik KD, Broadhurst RJ, Stewart-Wynne EG. Predicting survival for 1 year among different subtypes of stroke—results from the Perth Community Stroke Study. **Stroke** 1994; 25: 1935–1944.
45. Davila-Roman CG, Barzilai B, Wareing TH, Murphy SF, Schechtman KB, Kouchoukos NT. Atherosclerosis of the ascending aorta prevalence and role as an independent predictor of cerebrovascular events in cardiac patients. **Stroke** 1994; 25: 2010–2015.
46. Steimle AE, Stevenson LW, Fonarow GC, Hamilton MA, Moriguchi JD. Prediction of improvement in recent onset cardiomyopathy after referral for heart transplantation. **J Am Coll Cardiol** 1994; 23: 553–559.
47. Kaye DM, Lambert GW, Lefkovits J, Morris M, Jennings G, Esler MD. Neurochemical evidence of cardiac sympathetic activation and increased central nervous system norepinephrine turnover in severe congestive heart failure. **J Am Coll Cardiol** 1994; 23: 570–578.
48. Foley JB, Brown RIG, Penn IM. Thrombosis and restenosis after stenting in failed angioplasty: Comparison with elective stenting. **Am Heart J** 1994; 128: 12–20.
49. Pelliccia F, d'Amati G, Cianfrocca C, *et al.* Histomorphometric features predict 1-year outcome of patients with idiopathic dilated cardiomyopathy considered to be at low priority for cardiac transplantation. **Am Heart J** 1994; 128: 316–325.
50. Barefoot JC, Patterson JC, Haney TL, Cayton TG, Hickman JR Jr, Williams RB. Hostility in asymptomatic men with angiographically confirmed coronary artery disease. **Am J Cardiol** 1994; 74: 439–442.
51. Younis L, Stratmann H, Takase B, Byers S, Chaitman BR, Miller DD. Preoperative clinical assessment and dipyridamole thallium-201 scintigraphy for prediction and prevention of cardiac events in patients having major noncardiovascular surgery and known or suspected coronary artery disease. **Am J Cardiol** 1994; 74: 311–317.
52. Morris JJ, Tan YS. Autotransfusion: Is there a benefit in a current practice of aggressive blood conservation? **Ann Thorac Surg** 1994; 58: 502–508.
53. Morris JJ, Schaff HV, Mullany CJ, *et al.* Determinants of survival and recovery of left ventricular function after aortic valve replacement. **Ann Thorac Surg** 1993; 56: 22–30.
54. Lansman SL, Cohen M, Galla JD, *et al.* Coronary bypass with ejection fraction of 0.20 or less using centigrade cardioplegia: Long-term follow-up. **Ann Thorac Surg** 1993; 56: 480–486.
55. Ottino G, Bergerone S, Leo MD, *et al.* Aortocoronary bypass results: A discriminant multivariate analysis of risk factors of operative mortality. **J Cardiovasc Surg** 1990; 31: 20–25.
56. He GW, Acuff TE, Ryan WH, *et al.* Determinants of operative mortality in elderly patients undergoing coronary artery bypass grafting. **J Thorac Cardiovasc Surg** 1994; 108: 73–81.
57. Curtis JJ, Walls JT, Salam NH, *et al.* Impact of unstable angina on operative mortality with coronary revascularization at varying time intervals after myocardial infarction. **J Thorac Cardiovasc Surg** 1991; 102: 867–873.
58. Hosmer DW Jr, Lemeshow S. Applied Logistic Regression. John Wiley & Sons, New York, 1989.
59. Phillips NR, Waters D, Havel RJ. Plasma lipoproteins and progression of coronary artery disease evaluated by angiography and clinical events. **Circulation** 1993; 88: 2762–2770.
60. Weintraub WS, Craver JM, Cohen CL, Jones EL, Guyton RA. Influence of age on results of coronary artery surgery. **Circulation** 1991; 84(Suppl. III): 226–235.

61. Mullany CJ, Darling GE, Pluth JR, *et al.* Early and late results after isolated coronary artery bypass surgery in 159 patients aged 80 years and older. **Circulation** 1990; 82(Suppl. IV): 229–236.
62. Draper NR, Smith H. Applied Regression Analysis, 2nd Ed. John Wiley & Sons, New York, 1981.
63. Mantel N. Why step-down procedures in variable selection. **Technometrics** 1970; 12: 621–625.
64. Dobson AJ. Introduction to Statistical Modelling. Chapman & Hall, London, 1983.
65. Neter J, Wasserman W, Kutner MH. Applied Linear Statistical Models, 2nd Ed. Richard D. Irwin, Homewood, Illinois, 1985.
66. Afifi AA, Clark V. Computer-Aided Multivariate Analysis. Lifetime Learning Publications, Belmont, California, 1984.
67. Weisberg S. Applied Linear Regression, 2nd Ed. John Wiley & Sons, New York, 1985.
68. Rothman KJ. Modern Epidemiology. Little, Brown and Company, Boston, 1986.
69. SAS Institute, Inc. SAS/STAT User's Guide, version 6, 4th Ed., Vol. 2. SAS Institute, Inc., Cary, North Carolina, 1993.
70. Altman DG, Andersen PK. Bootstrap investigation of the stability of a Cox Regression Model. **Stat Med** 1989; 8: 771–783.
71. Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. **Br J Math Stat Psychol** 1992; 45: 265–282.
72. Flack VF, Chang PC. Frequency of selecting noise variables in subset regression analysis: A simulation study. **Am Stat** 1987; 41: 84–86.
73. Little RJA, Rubin DB. Statistical Analysis with Missing Data. John Wiley & Sons, New York, 1987.
74. Bendel RB, Afifi AA. Comparison of stopping rules in forward "stepwise" regression. **J Am Stat Assoc** 1977; 72: 46–53.
75. Costanza MC, Afifi AA. Comparison of stopping rules in forward stepwise discriminant analysis. **J Am Stat Assoc** 1979; 74: 777–785.
76. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. **Am J Epidemiol** 1989; 129: 125–137.
77. Simpson EH. The interpretation of interaction in contingency tables. **J R Stat Soc B** 1951; 13: 238–241.
78. Kendall M. Multivariate Analysis, 2nd Ed. Macmillan, New York, 1980.
79. Kraus JF, Anderson CL, Hemyari P, Zador PL, Sun GW. Features of fatal- and severe-injury motorcycle crashes according to vehicle design type, California 1985. **J Traffic Med** 1992; 20: 5–13.
80. Huberty CJ. Problems with stepwise methods—better alternatives. **Adv Social Sci Methodol** 1989; 1: 43–70.