J Clin Epidemiol Vol. 49, No. 12, pp. 1373–1379, 1996

S0895-4356/96/$15.00
PII S0895-4356(96)00236-3

A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis

Peter Peduzzi,[1,4,*] *John Concato,*[2,3] *Elizabeth Kemper,*[1,4] *Theodore R. Holford,*[4] *and Alvan R. Feinstein*[2,3,4]

[1]COOPERATIVE STUDIES PROGRAM COORDINATING CENTER AND THE [2]MEDICAL SERVICE, VETERANS AFFAIRS MEDICAL CENTER, WEST HAVEN CONNECTICUT 06516; AND THE DEPARTMENTS OF [3]MEDICINE (CLINICAL EPIDEMIOLOGY UNIT) AND [4]EPIDEMIOLOGY AND PUBLIC HEALTH, YALE UNIVERSITY SCHOOL OF MEDICINE, NEW HAVEN, CONNECTICUT 06510

ABSTRACT. We performed a Monte Carlo study to evaluate the effect of the number of events per variable (EPV) analyzed in logistic regression analysis. The simulations were based on data from a cardiac trial of 673 patients in which 252 deaths occurred and seven variables were cogent predictors of mortality; the number of events per predictive variable was (252/7=) 36 for the full sample. For the simulations, at values of EPV = 2, 5, 10, 15, 20, and 25, we randomly generated 500 samples of the 673 patients, chosen with replacement, according to a logistic model derived from the full sample. Simulation results for the regression coefficients for each variable in each group of 500 samples were compared for bias, precision, and significance testing against the results of the model fitted to the original sample.

For EPV values of 10 or greater, no major problems occurred. For EPV values less than 10, however, the regression coefficients were biased in both positive and negative directions; the large sample variance estimates from the logistic model both overestimated and underestimated the sample variance of the regression coefficients; the 90% confidence limits about the estimated values did not have proper coverage; the Wald statistic was conservative under the null hypothesis; and paradoxical associations (significance in the wrong direction) were increased. Although other factors (such as the total number of events, or sample size) may influence the validity of the logistic model, our findings indicate that low EPV can lead to major problems. *Copyright © 1996 Elsevier Science Inc.* J CLIN EPIDEMIOL 49;12:1373–1379, 1996.

KEY WORDS. Monte Carlo, bias, precision, significance testing

INTRODUCTION

Multivariable methods of analysis have been suspected of producing problematic results if too few outcome events are available relative to the number of independent variables analyzed in the model [1]. The main concerns have been accuracy and precision of the regression coefficients, and potentially misleading associations. Three types of errors have been discussed: overfitting (Type I error) occurs when too many variables, some of which may be "noise," are selected for retention in the final model; underfitting (Type II error) occurs when important variables are omitted from the final model; and paradoxical fitting (Type III error) is produced when a particular factor is given an incorrect direction of association which is the opposite of the true effect.

Because of these problems, general guidelines have been suggested for the minimum number of events per variable (EPV) required in multivariate analysis. On theoretical grounds, Harrell and colleagues [2] advocated a criterion equivalent to a minimum of 10–20 EPV. In a simulation study of forward stepwise multiple linear regression, Freedman and Pee [3] demonstrated that the Type I error was inflated when the ratio of the number of variables to the number of observations was greater than 0.25, corresponding to an EPV < 4. In simulation studies of the effect of EPV on proportional hazards regression [4,5], we recently suggested that at least 10 events per variable analyzed were desirable to maintain the validity of the model.

Because the impact of EPV may not be the same for all multivariable methods, we conducted a Monte Carlo study for the effect in logistic regression analysis. For the logistic model, the number of outcome events is the smaller number of binary outcomes (e.g., alive versus dead). Thus, a particular study may have many subjects, but too few deaths for a valid analysis. To investigate this problem, we conducted simulations using data from a cardiac trial having 252 deaths (events) among 673 patients. Seven known prognostic variables were selected for analysis, yielding an EPV of 252/7=) 36 for the full sample. The simulations were conducted for selected values of EPV ranging from 2 to 25. Results were compared with the model fitted to the original sample to examine bias, precision, and significance testing of the regression coefficients.

METHODS

Design of Simulation Study

Using sampling with replacement, 500 hundred simulations were each conducted at individual settings of EPV = 2, 5, 10, 15, 20, and 25. Deaths and survivors were separately sampled based on the predicted probability of dying (P_i) or surviving ($Q_i = 1 - P_i$) by the logistic model, where $P_i = 1/\{1 + \exp[-(\alpha + \mathbf{X}_i\beta)]\}$; α is the intercept term; $\mathbf{X}_i = (X_{i1}, \ldots, X_{i7})$ is the set of covariate values for patient i; and $\beta = (\beta_1, \ldots, \beta_7)$, is the set of corresponding values

*Address for correspondence: Peter Peduzzi, Cooperative Studies Program (151A), VA Medical Center, 950 Campbell Avenue, West Haven, Connecticut 06516.

Accepted for publication on 13 May 1996.

of the regression coefficients estimated from the full sample with EPV = 36. We wanted to draw a sample in which the associations with the covariates, β, was specified, while the number of events was fixed in the simulation. This goal was accomplished by selecting patient i, with probability $S_i(\text{death}) = P_i/\Sigma_{\{j=1,673\}}P_j$, if we were selecting a death, and $S_i(\text{survival}) = Q_i/\Sigma_{\{j=1,673\}}Q_j$ for a survivor.

The algorithm for selecting deaths was based on cumulative death and survival selection probabilities, $C_i(\text{death}) = \Sigma_{\{k=1,i\}}S_k(\text{death})$ and $C_i(\text{survival}) = \Sigma_{\{k=1,i\}}S_k(\text{survival})$, respectively; and a generated set of uniform random numbers, $\mathbf{u} = u_1, \ldots, u_j$, between 0 and 1, where $j = 7 \times \text{EPV}$. Patient i (1 to 673) was selected as a death if $C_{i-1}(\text{death}) < u \leq C_i(\text{death})$. Selection of deaths continued, with replacement, until the required number of $7 \times \text{EPV}$ deaths was obtained. For example, 2 EPV required $7 \times 2 = 14$ deaths. Selection of the $(673 - 7 \times \text{EPV})$ survivors was done in a similar manner using the cumulative survival scores, $C_i(\text{survival})$, and a generated set of $(673 - 7 \times \text{EPV})$ uniform random numbers. Thus, for each EPV simulation, the same subject could be selected more than once in a sample both as a death and as a survivor. The allocation of deaths and survivors in a simulation was independent of the actual outcome for each subject; that is, a subject who died in the study population could be selected as a survivor and vice versa.

At each value of EPV, the process of selecting deaths and survivors was repeated 500 times, generating 500 samples of 673 patients. For each sample, a logistic regression model was fitted and the resulting coefficients and their variances were saved in a data set for the analyses described later. The simulations were performed in SAS using an IBM RISC computer. A criterion of 10^{-6} was used for convergence of the maximum likelihood estimates. Simulations in which convergence was not obtained were excluded and not replaced.

The cited method of (retrospective) sampling was used to vary EPV, while at the same time leaving the regression parameters, β, unchanged. This strategy can be justified by using well-known properties of log-linear models and multinomial sampling. By conditioning on the deaths, the score for patient i represents the probability of observing the covariate pattern $\mathbf{X}_i$, or $P(\mathbf{X}_i|\text{death}) = P_i/\Sigma_{\{j=1,673\}}P_j = K_1P_i$, and by conditioning on survivors we obtain $P(\mathbf{X}_i|\text{survival}) = Q_i/\Sigma_{\{j=1,673\}}Q_j = K_2Q_i$. The ratio of the death to survivor score is proportional to $\exp(\alpha^* + \mathbf{X}_i\beta)$, which is a log-linear model where $\alpha^* = \alpha + \ln(K_1/K_2)$. If we apply Bayes theorem, we obtain the probability of death given the covariate pattern (as in prospective sampling):

$$
\begin{aligned}
&P(\text{death}|\mathbf{X}_i)\\
&= P(\text{death})P(\mathbf{X}_i|\text{death})/\{P(\text{death})P(\mathbf{X}_i|\text{death})\\
&\qquad + P(\text{survival})P(\mathbf{X}_i|\text{survival})\}\\
&= 1/\{1 + \exp[-\ln\{K_1P(\text{death})/K_2P(\text{survival})\} - \alpha - \mathbf{X}_i\beta]\}\\
&= 1/\{1 + \exp(-\alpha' - \mathbf{X}_i\beta)\},
\end{aligned}
$$

where

$$\alpha' = \ln\{K_1P(\text{death})/K_2\,P/(\text{survival})\} + \alpha.$$

In a similar manner, we can show that $P(\text{survival}|\mathbf{X}_i)$ is equal to $1/\{1 + \exp(\alpha' + \mathbf{X}_i\beta)\}$, and is consistent with the logistic model. Thus, by conditioning on the outcome (death or survival), sampling retrospectively will yield the same estimates for β as sampling prospectively conditional on the covariate patterns [6,7].

Retrospective sampling gives the distribution of estimates of the regression coefficients in relation to the observed EPV, while prospective sampling considers estimates in relation to the expected EPV. Our goal was to examine the behavior of the logistic model with respect to the observed number of events per variable, which is the quantity known in practice, and we follow the same strategy used in our previous investigation of the proportional hazards model [4,5]. The results of simulations using prospective sampling were virtually identical to those obtained for retrospective sampling reported here.

Statistics

The simulation results of EPV of 2, 5, 10, 15, 20, and 25 were evaluated relative to the model fitted to the original (full) sample with EPV = 36, using previously described measures of accuracy, precision, and significance testing [4]. We also examined the distribution of the regression coefficients and tested for normality using the Kolmogorov D statistic [8]. All statistics were computed conditional on obtaining convergence of the logistic regression model. Samples in which the model did not converge were excluded from analysis. Although these samples provide some information about the parameter estimates (i.e., upper or lower bounds), they do not provide useful estimates of the effect of covariates.

Accuracy of coefficients was assessed by calculating the average percent relative bias for each of the $k = 1, \ldots, 7$ regression coefficients, and each of the $m = 1, \ldots, M$ simulations that converged, as $100^*\Sigma_{\{m=1,M\}}(\beta_{km} - \beta_{k,\text{true}})/(M^*\beta_{k,\text{true}})$ where $\beta_{k,\text{true}}$ was the "true" value of the coefficient obtained from the full sample. Another measure of accuracy was the proportion of simulations in which the bias exceeded $\pm$ 100%.

Precision of coefficients was determined by calculating and then comparing the "sample" and "model" variance of each regression coefficient. The sample variance was calculated in the usual manner as $\Sigma_{\{m=1,M\}}(\beta_{km} - \bar{\beta}_k)^2/(M - 1)$. Model variance was determined as the average of the variances from the logistic model for each coefficient over all M models that converged, that is, $\Sigma_{\{m=1,M\}}\text{Var}(\beta_{km})/M$. The ratio of model to sample variance was used to assess the large sample properties of the model; ratios different from a value of one indicated that these properties may not hold.

The statistical significance of the regression coefficients was evaluated in four ways. First, the coverage of 90% confidence intervals was determined as the proportion of simulations in which the 90% confidence interval about the estimated coefficient included the true value. Second, power was calculated as the proportion of simulations in which the coefficient divided by its standard error (Z-statistic) exceeded normal deviate for a one-sided significance test at the 10% level (1.28). Third, the proportion of simulations in which the Z-statistic was less than -1.28 would indicate the chance of observing significance in the wrong direction, described as "paradoxical fitting" [4] or Type III error [9]. Fourth, to assess the validity of the Z-statistic, we conducted simulations under the null hypothesis, by setting all the regression coefficients simultaneously equal to zero ($\beta = 0$) and using the same simulation strategy described above. The distribution of the Z-statistics was then examined and Type 1 error evaluated.

DATA

The simulations used data from the Department of Veterans Affairs Cooperative Study of Coronary Artery Surgery [10]. In this study, 686 patients with stable angina pectoris and angiographically proven coronary artery disease were enrolled between the years 1972 and 1974 and followed thereafter for a minimum of 10 years. Seven

TABLE 1. Summary statistics of baseline risk factors in original complete group

Factor	Prevalence	Multivariable logistic regression estimates			
		Coefficient	Standard error	Wald *p* value	Odds ratio
Intercept		−1.86	0.24	<0.01	
ST depression	0.25	0.46	0.19	0.02	1.59
History of hypertension	0.29	0.52	0.19	<0.01	1.67
NYHA Class III or IV	0.59	0.28	0.17	0.11	1.32
History of CHF	0.07	0.51	0.33	0.12	1.67
History of diabetes	0.13	0.56	0.25	0.02	1.75
Number of vessels diseased					
1	0.14	0.33	0.12	<0.01	1.39
2	0.32				
3	0.54				
Abnormal left ventricular contractility	0.50	0.61	0.17	<0.01	1.85

Global chi-square score statistic with 7 degrees of freedom = 56.7, $p < 0.01$.

known predictors of survival (with bivariate *p* values (2-sided) <0.10 for each variable) were selected for the simulations: ST segment depression (STD) on the resting baseline ECG, history of hypertension (HTN), New York Heart Association Functional Class III or IV versus I or II (NYHA), history of congestive heart failure (CHF), history of diabetes mellitus (DM), number of coronary arteries with significant lesions (VES), and the presence of a left ventricular contraction abnormality (LVC). In the current analyses, all variables were coded as 1 for presence of the factor and 0 for absence of the factor, except the number of diseased vessels, coded as 0 for 1-vessel disease, 1 for 2-vessel disease, and 2 for 3-vessel disease. Complete data for the 7 variables were available in 673 patients, of whom 252 died during the first 10 years of follow-up, yielding an EPV of (252/7=) 36 for the full sample. Table 1 summarizes the results of the multivariate logistic regression model applied to the full sample.

RESULTS

Although all 500 sample models converged for EPV ≥ 10 and 497 models converged for EPV = 5, only 377 (77%) models converged for EPV = 2. Thus, the logistic model did not always achieve convergence at low EPV.

Figure 1 shows the effect of EPV on the frequency distribution of the values of the regression coefficients for the variable CHF. As EPV decreased, the distributions became more dispersed and "flat-

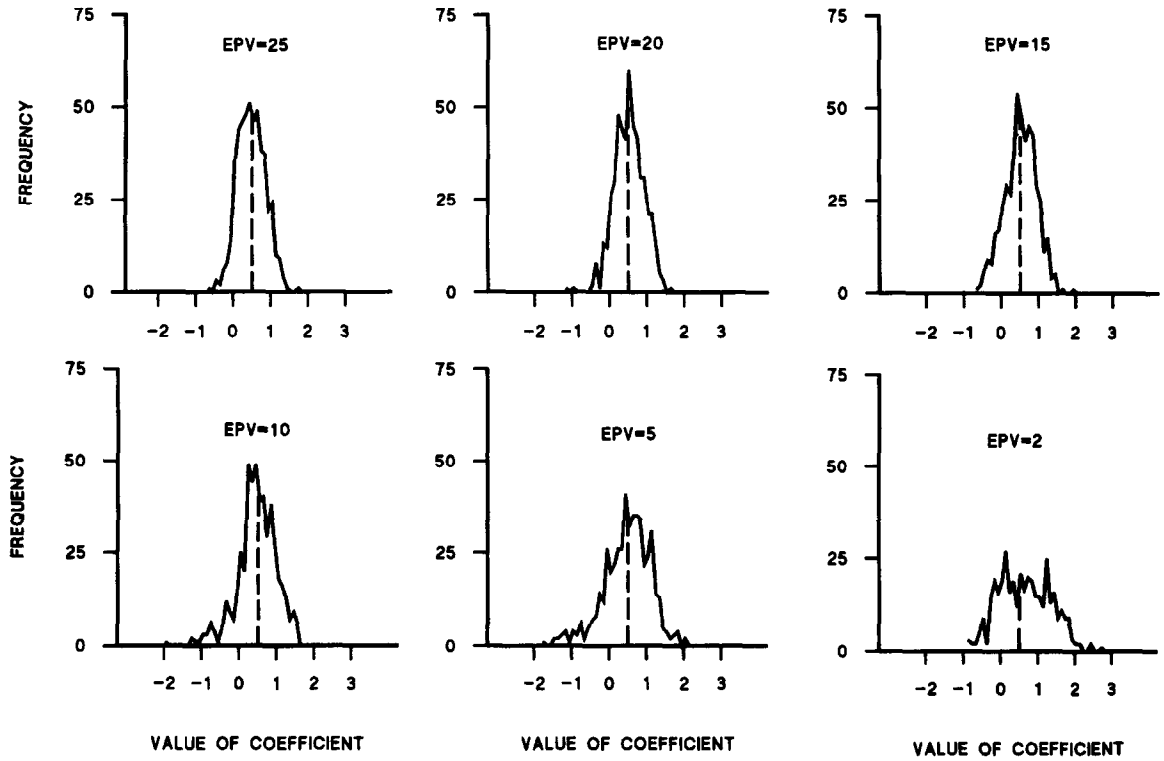

FIGURE 1. Number of events per variable, and frequency distribution of estimated regression coefficients for congestive heart failure (CHF). The vertical (*dashed*) line is the true value for the regression coefficient.

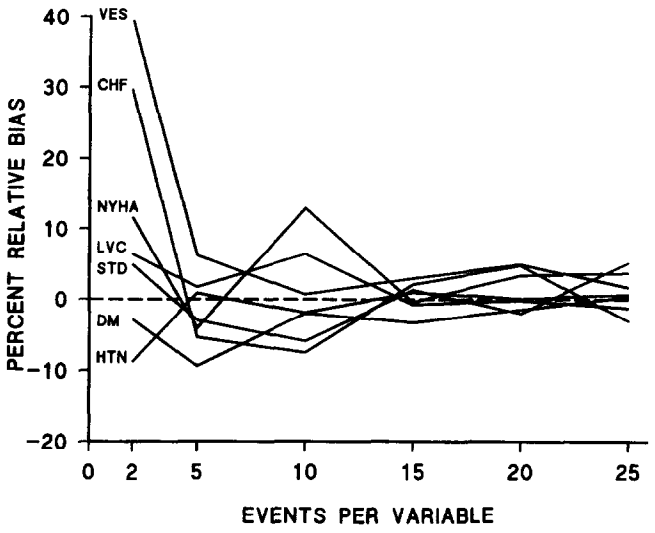

FIGURE 2. Number of events per variable, and average percent relative bias. Abbreviations for variables are: VES = number of coronary arteries with significant lesions; CHF = history of congestive heart failure; NYHA = New York Heart Association Functional Class III or IV; LVC = presence of a left ventricular contraction abnormality; STD = ST depression on the resting baseline electrocardiogram; DM = history of diabetes mellitus; and HTN = history of hypertension.

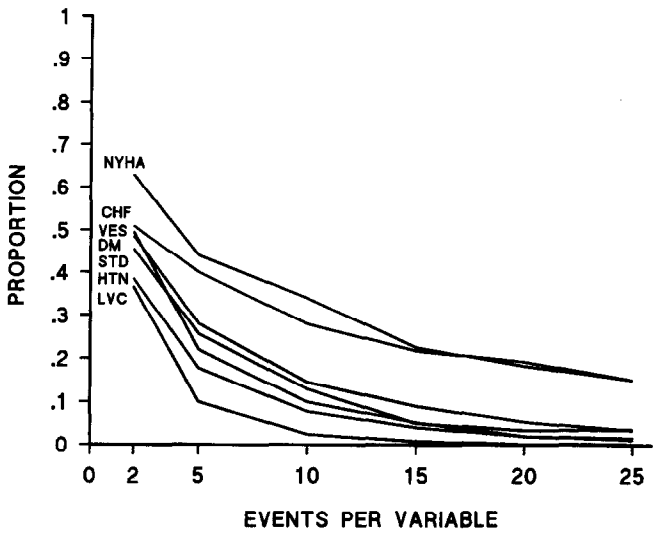

FIGURE 3. Number of events per variable, and proportion of simulations in which percent relative bias exceeded ± 100%. Abbreviations are as indicated in Figure 2.

ter," particularly for EPV < 10. For example, the minimum and maximum values of the regression coefficient were −0.67 and 1.71 at 25 EPV compared with −0.85 and 2.75 at 2 EPV. Thus, inaccurate estimation of the actual regression coefficient value was more likely at low EPV. Similar patterns were observed for the other 6 variables (data not shown).

The normality of the distribution of the regression coefficients was tested with the Kolmogorov D statistic (data not shown). Departures from normality ($p < 0.05$), with "flattened" distributions and long "tails" in both directions, were more common as EPV decreased. The frequency of "not normal" distributions was 7 at 2 EPV, 3 at 5 EPV, 2 at 10 EPV, 1 at both 15 EPV and 20 EPV, and 0 at 25 EPV.

Percent relative bias of coefficients, as graphically displayed in Figure 2, increased with decreasing EPV. The average bias was within ± 10% of the true value for EPV > 10. At 2 EPV, however, the bias increased dramatically, so that regression coefficients were overestimated by an average of 30% for CHF and by 40% for VES. This problem is further described in Figure 3, which shows the proportion of simulations in which the absolute error exceeded ± 100%. The proportions increased substantially below 10 EPV, and exceeded 0.35 for all factors at 2 EPV. The proportions for NYHA and CHF were still greater than 0.20 at 25 EPV, apparently due to the relatively small impact of NYHA on outcome, and the relatively large standard error for CHF which arises because of low prevalence.

The sample variance of the regression coefficients, displayed in Figure 4, showed the expected increase as the EPV decreased. The corresponding effect on model variance is shown in Figure 5 for the ratio of model to sample variance. Marked departures from the expected value of 1 occurred at 2 EPV, where the ratio was overestimated by nearly 100% for CHF and by 25% for DM, and underestimated by 20% for VES and LVC and by 10% for STD and HTN.

Figure 6 shows the proportion of simulations in which the 90% confidence limits about the estimated value included the true value. Both overcoverage and undercoverage occurred, with greatest variability in coverage at low EPV. The median proportion of coverage among the 7 variables was 0.92 at 2 EPV, 0.91 at 5 EPV and 0.90 for EPV ≥ 10.

Figure 7 shows the proportion of simulations in which the Z-statistic (square root of Wald statistic) exceeded the standard normal deviate of 1.28 which gives the power for a 10% significance test. The power for all variables decreased slowly with decreasing number of events up to 10 EPV, and thereafter dropped sharply. The proportion of simulations in which the Z-statistic was paradoxically reversed to values less than −1.28 was low (data not shown). For four of the factors the proportions for paradoxical associations increase

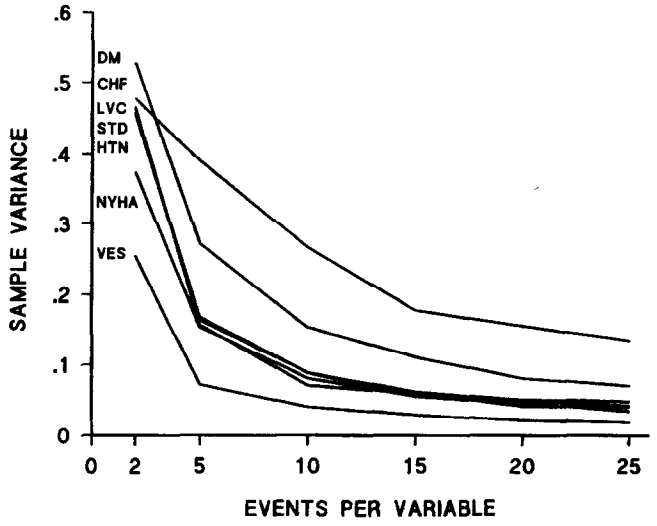

FIGURE 4. Number of events per variable, and sample variance of the estimated regression coefficients. Abbreviations are as indicated in Figure 2.

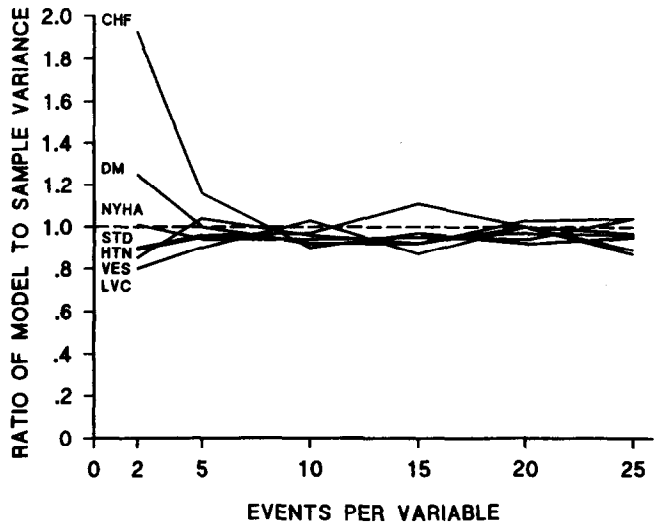

FIGURE 5. Number of events per variable, and ratio of model to sample variance. Abbreviations are as indicated in Figure 2.

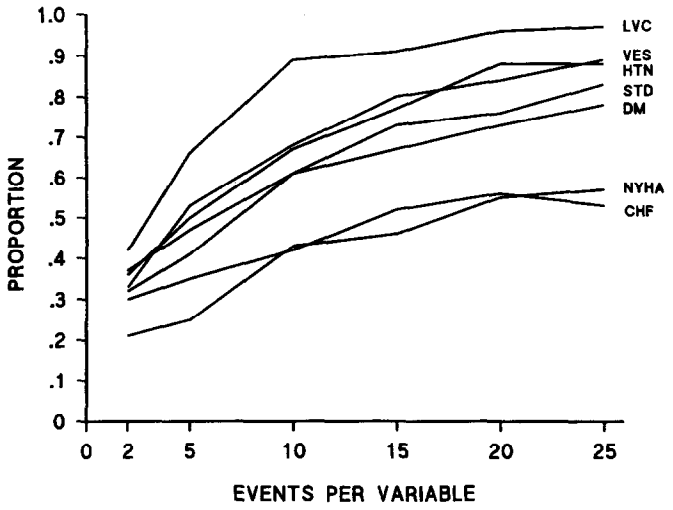

FIGURE 7. Proportion of simulations in which the Z-statistic (coefficient/standard error) Exceeded the standard normal deviate of 1.28 for 90% power. Abbreviations are as indicated in Figure 2.

at low EPV: at 2 EPV, the proportions were 0.03 for NYHA, 0.02 for STD and HTN, and 0.01 for VES.

The normality of the distributions of the Z-statistic were tested under the null hypothesis of no covariate effects ($\beta = 0$). Departures from normality were common below 10 EPV; the distributions were skewed to the left. At 5 EPV the distributions of four variables were not normal ($p < .05$) and the data suggest that none of the distributions was normal at 2 EPV. These phenomena are illustrated for one variable (CHF) in Figure 8. For example, the minimum and maximum values of the Z-statistic were −2.44 and 2.85 at 25 EPV compared with −3.48 and 0.88 at 2 EPV.

Finally, Table 2 displays the number of simulations that converged, and the overall percentage of occasions in which each variable was significant at the 10% level under the null hypothesis ($\beta = 0$) of no covariate effects. The global Type I error for each EPV was determined as the total number of variables found to be significant in all simulations divided by the total number of variables that were evaluated (i.e., seven times the number of simulations that converged). The global Type I error decreased from 10.3% for 25 EPV to 8.5% for 2 EPV, indicating that the Z-statistic became excessively conservative with decreasing EPV.

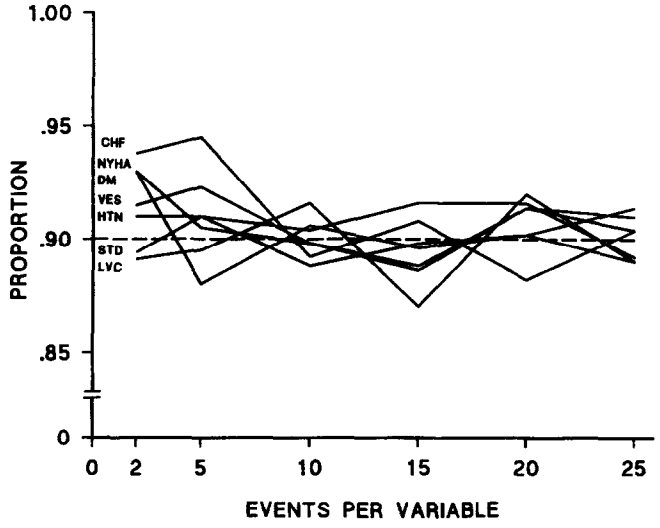

FIGURE 6. Proportion of Simulations in which the 90% confidence interval about the estimated regression coefficient included the true value. Abbreviations are as indicated in Figure 2.

DISCUSSION

These simulation studies demonstrate the problems that can occur when a logistic model contains few events relative to the number of independent variables being evaluated. As EPV decreased, the bias of the regression coefficients increased, often yielding extreme values for the maximum likelihood estimates. For example, at 2 EPV, more than one-third of the estimated regression coefficients were either twice as large or half as small as the true value. Even at 25 EPV, bias of the regression coefficient exceeded ± 100% in 20% of the estimates for CHF and NYHA. This finding cannot be explained solely by low prevalence of a variable: although CHF was uncommon (7%) in the population, NYHA had the highest prevalence (59%) of any variable. Similarly, the finding cannot be attributed only to low impact of a variable: although NYHA had low impact (odds ratio = 1.37), CHF had a greater impact (odds ratio = 1.67) than several other variables. A likely explanation is that either low prevalence or low impact can magnify the effects of a "too small" EPV.

The effect of decreasing EPV on estimates of power and variance was expected, because these two statistics are directly related to the number of events. At low EPV, however, an additional problem is that the large sample properties of the logistic model variance may not hold. It was overestimated (up to 100%) and underestimated (up to 20%), although a consistent pattern was not apparent. Paradoxical fitting (i.e., associations in the wrong direction) also showed an increased occurrence at low EPV, but the relative frequencies were small.

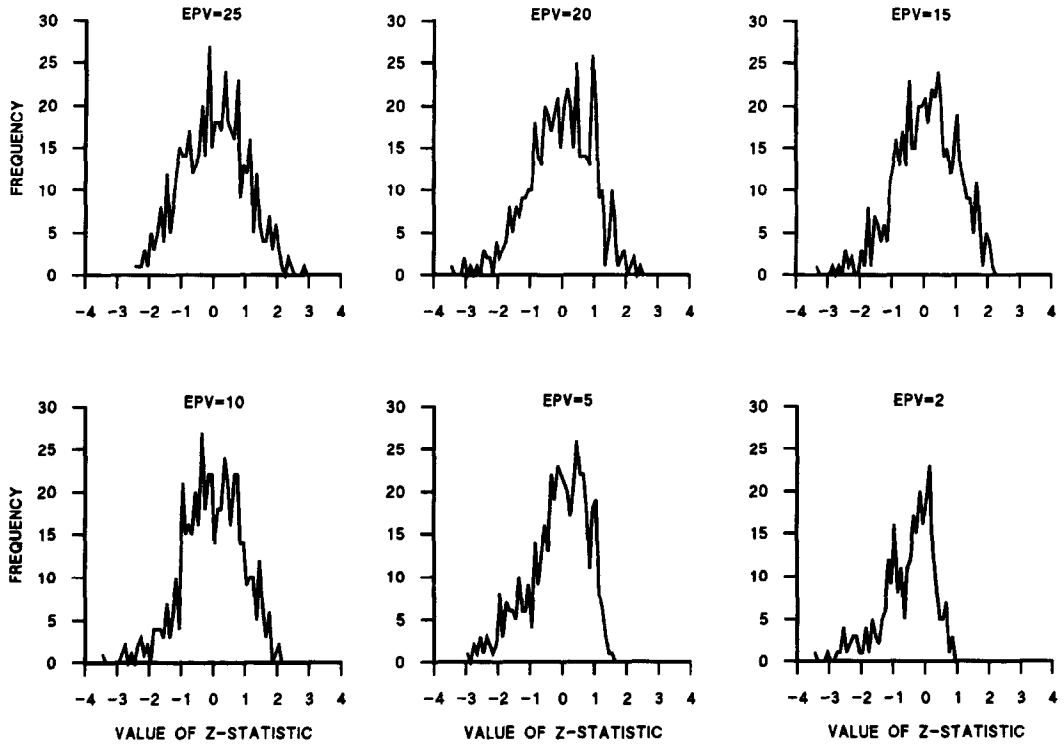

FIGURE 8. Distribution of the Z-statistic for congestive heart failure under the null hypothesis that the covariate has no effect with outcome.

TABLE 2. Number (*n*) and percent (%) of occasions in which variable was significant ($p < 0.10$) under null hypothesis of no covariate effects

	EPV = 25 (500)[a]		EPV = 20 (500)[a]		EPV = 15 (500)[a]		EPV = 10 (497)[a]		EPV = 5 (471)[a]		EPV = 2 (282)[a]	
Variable	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
STD	47	9.4	60	12.0	56	11.2	60	12.1	53	11.3	19	6.7
HTN	63	12.6	46	9.2	30	6.0	43	8.7	47	10.0	19	6.7
NYHA	51	10.2	57	11.4	61	12.2	61	12.3	50	10.6	28	9.9
CHF	50	10.0	45	9.0	45	9.0	32	6.4	37	7.9	29	10.2
DM	49	9.8	44	8.8	44	8.8	45	9.1	36	7.6	19	6.7
VES	48	9.6	47	9.4	41	8.2	52	10.5	41	8.7	25	8.9
LVC	51	10.2	48	9.4	59	11.8	53	10.7	48	10.2	29	10.3
Total[b]	359	10.3	346	9.9	336	9.6	346	9.9	312	9.5	168	8.5

Abbreviations are as indicated in Figure 2.

[a]Number of logistic regression analyses (in parentheses) that converged out of 500 samples.

[b]Total percent (or global *p* value) was calculated as the number of significant variables divided by total number of variables evaluated (=7 × number of samples that converged).

Another problematic finding was the effect of EPV on significance testing. The 90% confidence intervals about the simulated values exceeded 90% coverage for most variables at EPV ≤ 5, indicating that the intervals were too wide. For the simulations under the null hypothesis that the seven covariates had no relation with outcome, however, the Z-statistic was excessively conservative at low EPV, rejecting less frequently than the stated significance level. The simulations also offered strong evidence that the distribution of the Z-statistic was not normal and was skewed to the left at 2 EPV. Other problems with inference based on the Wald statistic in the current research have been previously noted in logistic regression analysis [11,12]. We did not, however, evaluate the effects of EPV for the score and likelihood ratio statistics.

An additional concern at low EPV involves the convergence of maximum likelihood estimates when the logistic regression coefficients are calculated. Convergence was obtained in less than 80% of the samples at 2 EPV. For the simulations under the null hypothesis, the convergence rate was only 60% at 2 EPV. The lower rate

of convergence may be attributed to variables with low prevalence rates, in that the chance of having the event when the risk factor was present was small under the null hypothesis. This problem may also occur in studies that evaluate variables with low prevalence rates and small effect sizes.

In an earlier simulation study of the logistic model [13], problems were noted with respect to the accuracy and precision of the regression estimates and associated tests of significance based on the Z-statistic. The simulations were based on a hypothetical population of 10,000 individuals and a single risk factor, and evaluated the effect of sample size on the validity from the model, but EPV was not examined. Regression coefficients were found to be both inaccurate and unreliable compared with the "true" population values, over a range of sample size from 1,000 to 7,500. Problems with the validity of the Z-statistic were observed under the null hypothesis, similar to the current study, in that the Type I error was either greater or less than the nominal value of 5%.

Our Monte Carlo study has several limitations. The simulations were based on the mortality rate in a single set of "real" data. The variables examined were all discrete and had only a moderate range of prevalence rates and associations (all positive) with outcome. No continuous measures or interactions between variables were included. Exact logistic regression [14] was not compared, and the performance of the likelihood ratio and score statistics was not evaluated. In addition, we checked one index (EPV) in our study, but others (such as the total number of events or sample size) may also affect the validity of the model.

In summary, the validity of the logistic model becomes problematic when the ratio of the numbers of events per variable analyzed becomes small. The parameter estimates may be biased and the usual tests of significance may not be valid. These results offer insight into problems of low EPV, and can guide the design of future Monte Carlo studies that include different indexes evaluated under a wider range of conditions.

This research was supported by the Department of Veterans Affairs Cooperative Studies Program of the Veteran's Health Administration. Dr. Concato is a Research Associate in the Career Development Program, Health Services Research and Development Service.

References

1. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. **Ann Int Med** 1993; 118: 201–210.
2. Harrel F, Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: Advantages, problems and suggested solutions. **Cancer Treatment Reports** 1985; 69: 1071–1077.
3. Freedman LS, Pee D. Return to a note on screening regression equations. **Am Statistician** 1989; 43: 279–282.
4. Concato J, Peduzzi P, Holford TR, Feinstein AR. The importance of events per independent variable (EPV) in proportional hazards analysis: I. Background, goals and general strategy. **J Clin Epidemiol** 1995; 48: 1495–1501.
5. Peduzzi P, Concato J, Feinstein AR, Holford TR. The importance of events per independent variable (EPV) in proportional hazards regression analysis: II. Accuracy and precision of regression estimates. **J Clin Epidemiol** 1995; 48: 1503–1510.
6. Breslow N, Powers W. Are there two logistic regressions for retrospective studies? **Biometrics** 1978; 34: 100–105.
7. Agresti A. **Categorical Data Analysis** New York: John Wiley; 1990.
8. Stephens MA. EDF statistics for goodness of fit and some comparisons. **J Am Stat Assoc** 1974; 69: 730–737.
9. Schwartz D, Flamant R, Lellouch J. (Translated by Healy, MJR.) Methods of comparison. **Clinical Trials** London: Academic Press; 1980: 77–86.
10. Peduzzi P, Detre K, Gage A. Veterans administration cooperative study of medical versus surgical treatment for stable angina—Progress report: Section 2—Design and baseline characteristics. **Prog Card Dis** 1985; 28: 235–243.
11. Hauck WW, Donner A. Wald's test as applied to hypotheses in logit analysis. **J Am Stat Assoc** 1977; 72: 851–853.
12. Jennings DE. Judging inference adequacy in logistic regression. **J Am Stat Assoc** 1986; 81: 471–476.
13. Lilienfeld D, Pyne DA. The logistic analysis of epidemiologic prospective studies: Investigation by simulation. **Stat Med** 1984; 3: 15–26.
14. Hirji KF, Mehta CR, Patel NR. Computing distributions for exact logistic regression. **J Am Stat Assoc** 1987; 72: 1110–1117.