American Journal of Epidemiology
Copyright © 2006 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved; printed in U.S.A.
Vol. 165, No. 6
DOI: 10.1093/aje/kwk052
Advance Access publication December 20, 2006

Original Contribution

Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression

Eric Vittinghoff and Charles E. McCulloch

From the Department of Epidemiology and Biostatistics, University of California, San Francisco, CA.

Received for publication March 15, 2006; accepted for publication August 15, 2006.

The rule of thumb that logistic and Cox models should be used with a minimum of 10 outcome events per predictor variable (EPV), based on two simulation studies, may be too conservative. The authors conducted a large simulation study of other influences on confidence interval coverage, type I error, relative bias, and other model performance measures. They found a range of circumstances in which coverage and bias were within acceptable levels despite less than 10 EPV, as well as other factors that were as influential as or more influential than EPV. They conclude that this rule can be relaxed, in particular for sensitivity analyses undertaken to demonstrate adequate control of confounding.

bias (epidemiology); coverage probability; event history analysis; model adequacy; type I error; variable selection

Abbreviation: EPV; events per predictor variable.

The rule of thumb that logistic and Cox models should be used with a minimum of 10 events per predictor variable (EPV) is based on two simulation studies (1–3). In these studies, only the numbers of events were varied; the sample size and the distribution and effects of the seven binary predictors were held constant at the values observed in a randomized trial (4). The results showed increasing bias and variability, unreliable confidence interval coverage, and problems with model convergence as EPV declined below 10 and especially below five, leading to the reasonable conclusion that results should be cautiously interpreted with less than 10 EPV.

Rules of thumb, such as 10 or more EPV, are useful signals for potential trouble and, for prediction, rules requiring 20 or more EPV may be appropriate (5). However, in analysis of causal influences in observational data, control of confounding may require adjustment for more covariates than the rule of 10 or more EPV allows (6). We carried out a simulation study to examine the influence of the factors not varied in the original studies, to identify circumstances where we might safely relax the rule of 10 or more EPV.

MATERIALS AND METHODS

We conducted a large factorial simulation study with binary as well as failure time endpoints, focusing on a primary predictor, either binary or continuous, and regarding the covariates as adjustment variables. We considered values of EPV from two to 16; models with a total of two, four, eight, and 16 predictor variables; sample sizes of 128, 256, 512, and 1,024; and values of β_1, the regression coefficient for the primary predictor, of 0, log(1.5), log(2), and log(4). The factorial omitted extreme cases with outcome prevalence of greater than 50 percent.

With a binary primary predictor, the other predictors were multivariate normal with pairwise correlation of 0.25. The binary primary predictor was generated with expected prevalence of 0.1, 0.25, or 0.5 and multiple correlation with the covariates of 0, 0.25, 0.5, or 0.75. With the continuous primary predictor, all predictors were multivariate normal and equally intercorrelated. The variance of the primary predictor was set to 0.16, for comparability with the binary primary predictors, and the multiple correlation between the primary predictor and adjustment variables was set to 0, 0.1,

Correspondence to Eric Vittinghoff, Box 0560, Department of Epidemiology and Biostatistics, University of California, 185 Berry Street, Suite 5700, San Francisco, CA 94107 (e-mail: eric@biostat.ucsf.edu).

Downloaded from http://aje.oxfordjournals.org/ at Kainan University on March 1, 2015

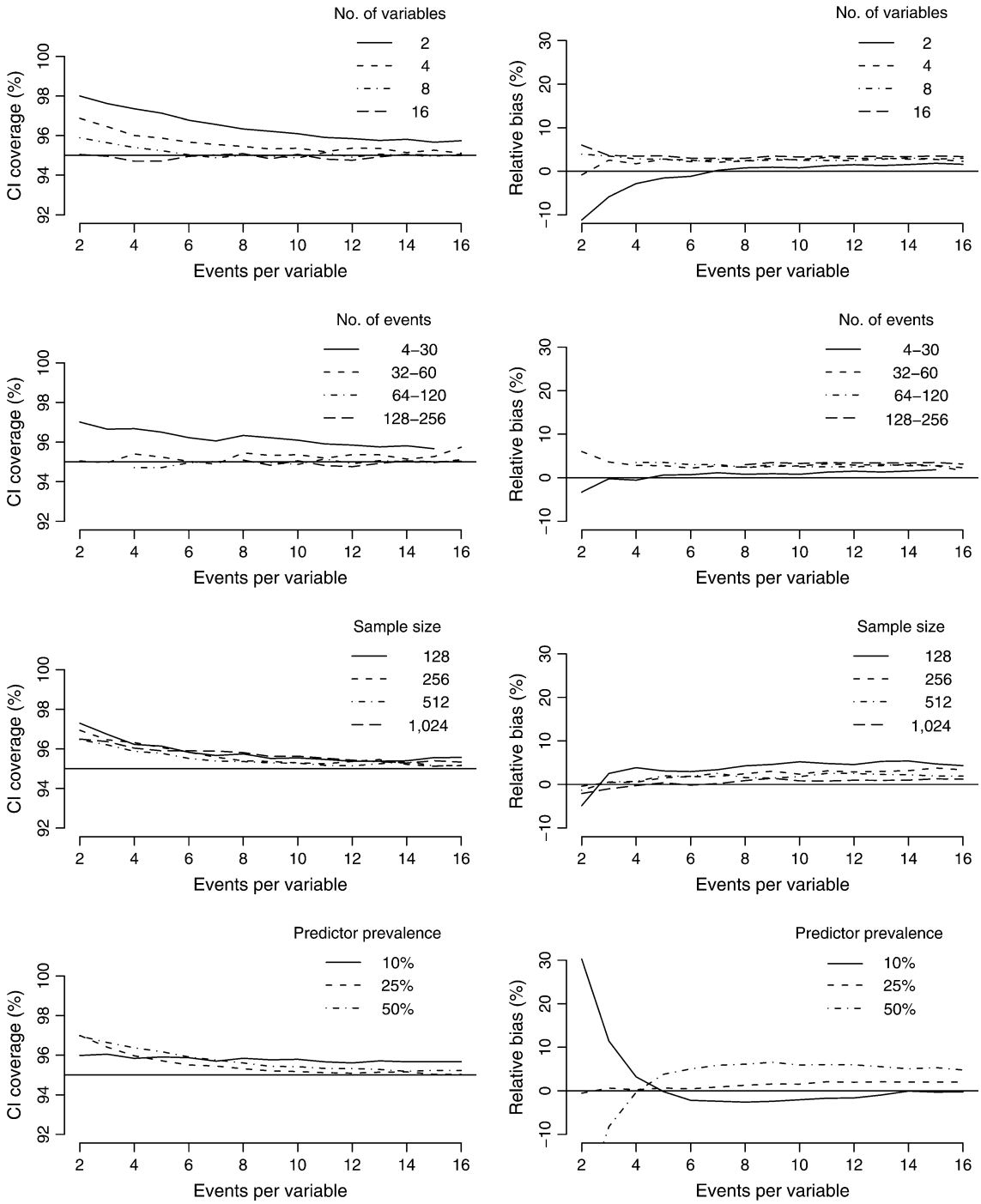

FIGURE 1. Logistic model with binary primary predictor. CI, confidence interval.

0.25, 0.5, or 0.9. The aggregate effect of the covariates was held constant across models with two, four, eight, and 16 predictors. We examined 9,328 and 3,392 scenarios with binary and continuous primary predictors, respectively.

In the logistic models, we kept the first "cases" and "controls" generated, up to the required numbers of each, taking advantage of the fact that under the logistic model only the intercept is affected by such retrospective sampling. For the Cox model, longer randomly generated failure times were censored after the required numbers of events had been "observed." For each combination of parameters, 500 data sets were generated and then analyzed

Downloaded from http://aje.oxfordjournals.org/ at Kainan University on March 1, 2015

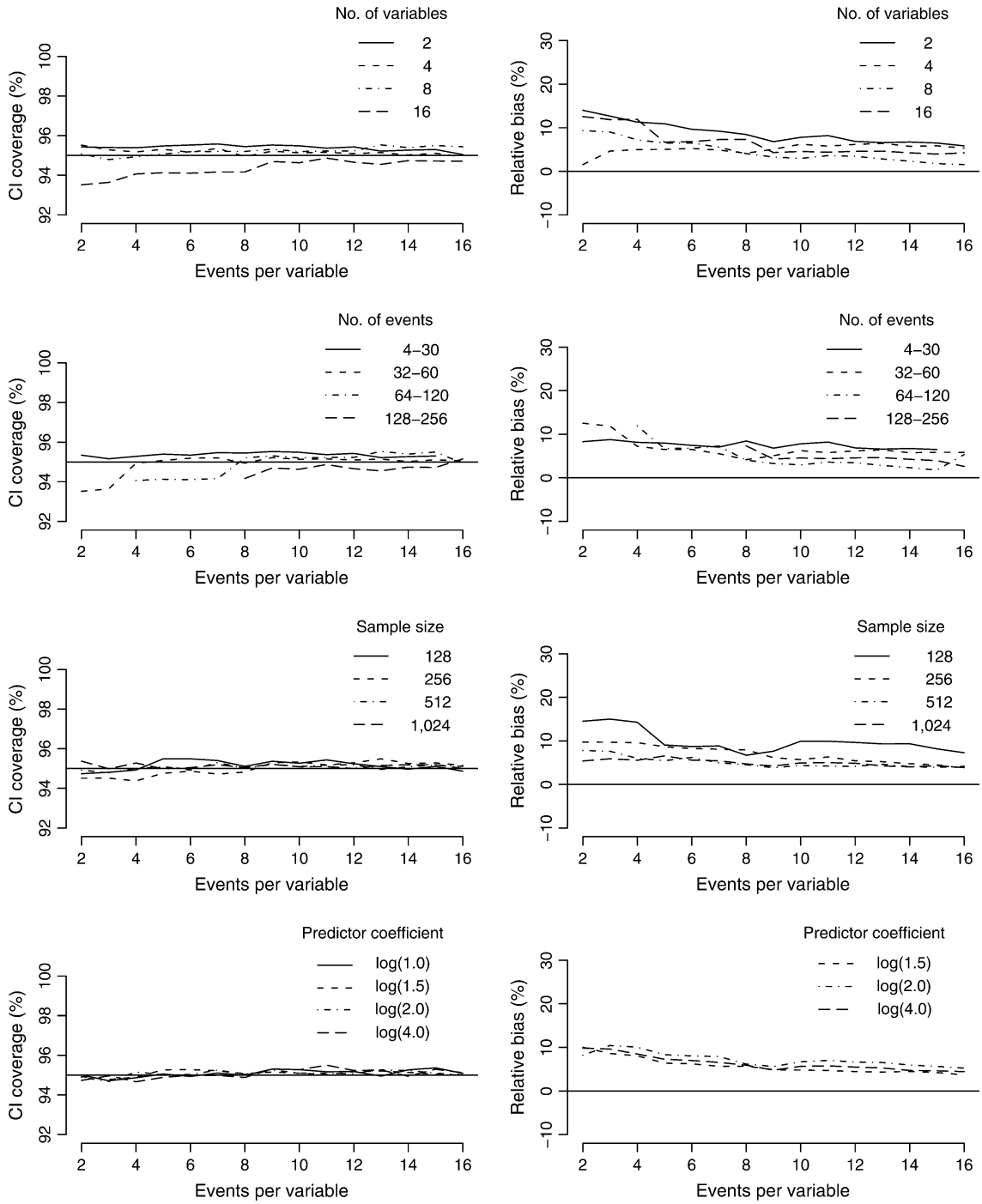

FIGURE 2. Logistic model with continuous primary predictor. CI, confidence interval.

in SAS, version 9.13, software (SAS Institute, Inc., Cary, North Carolina). Results from data sets for which the model did not converge were excluded from the computation of summary statistics.

Confidence interval coverage was estimated by the percentage of the retained data sets in which the Wald 95 percent confidence interval for β_1 included the true value. Relative bias was estimated for $\beta_1 > 0$ by the percentage difference between the average estimate and the true value. We also estimated the type I error rate or power of the two-sided Wald test of H_0 ($\beta_1 = 0$) by the proportion of data sets in which the test was statistically significant at $p < 0.05$.

Downloaded from http://aje.oxfordjournals.org/ at Kainan University on March 1, 2015

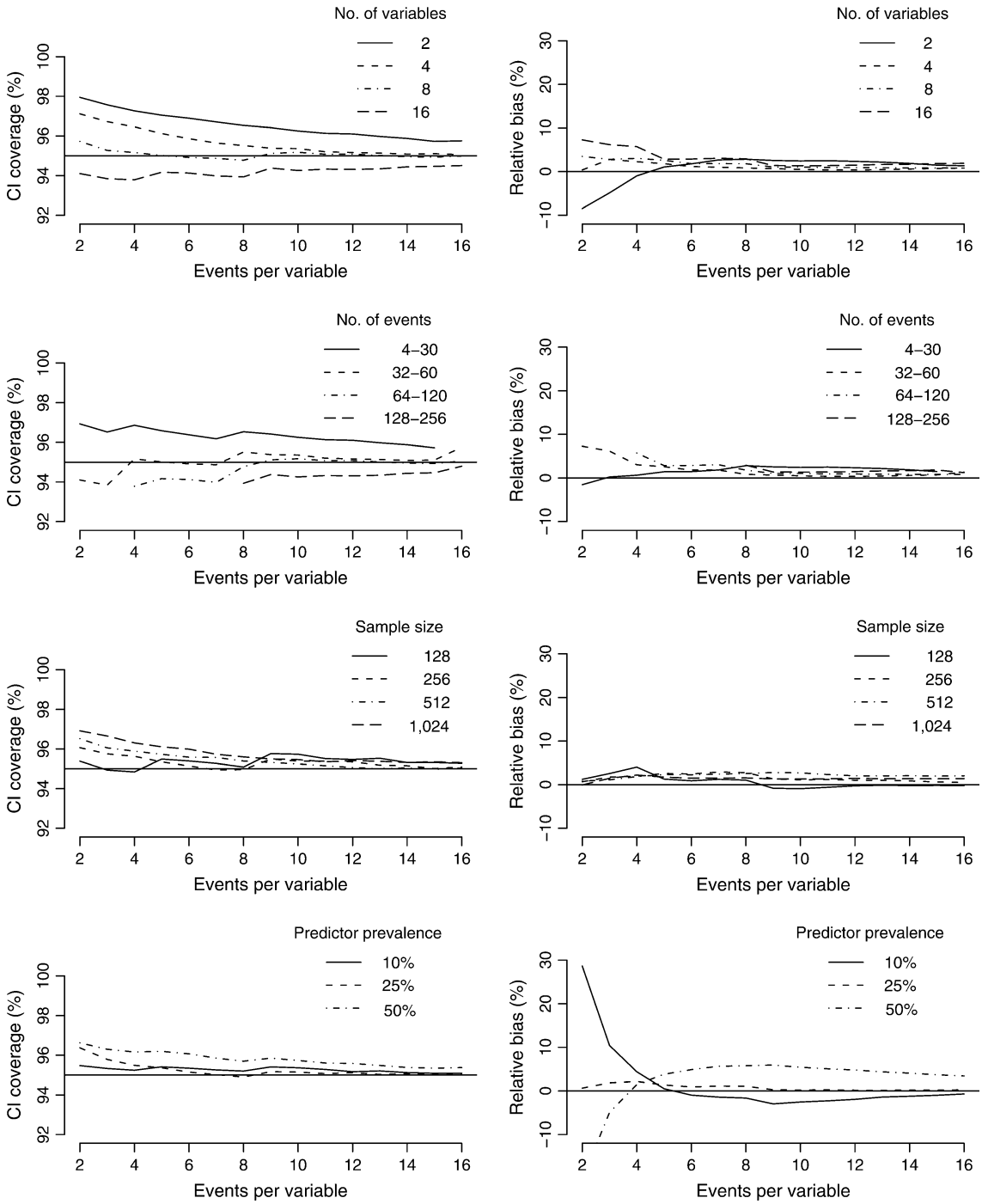

FIGURE 3. Cox model with binary primary predictor. CI, confidence interval.

Finally, we tabulated problematic scenarios with confidence interval coverage less than 93 percent, type I error rate greater than 7 percent, or relative bias greater than 15 percent, and we report the worst confidence interval coverage, type I error rate, and relative bias for each model and type of predictor.

RESULTS

Results are summarized in figures 1–4. The left column of each figure displays confidence interval coverage for β_1, and the right column shows relative bias. In each panel, average confidence interval coverage or relative bias is plotted for

Downloaded from http://aje.oxfordjournals.org/ at Kainan University on March 1, 2015

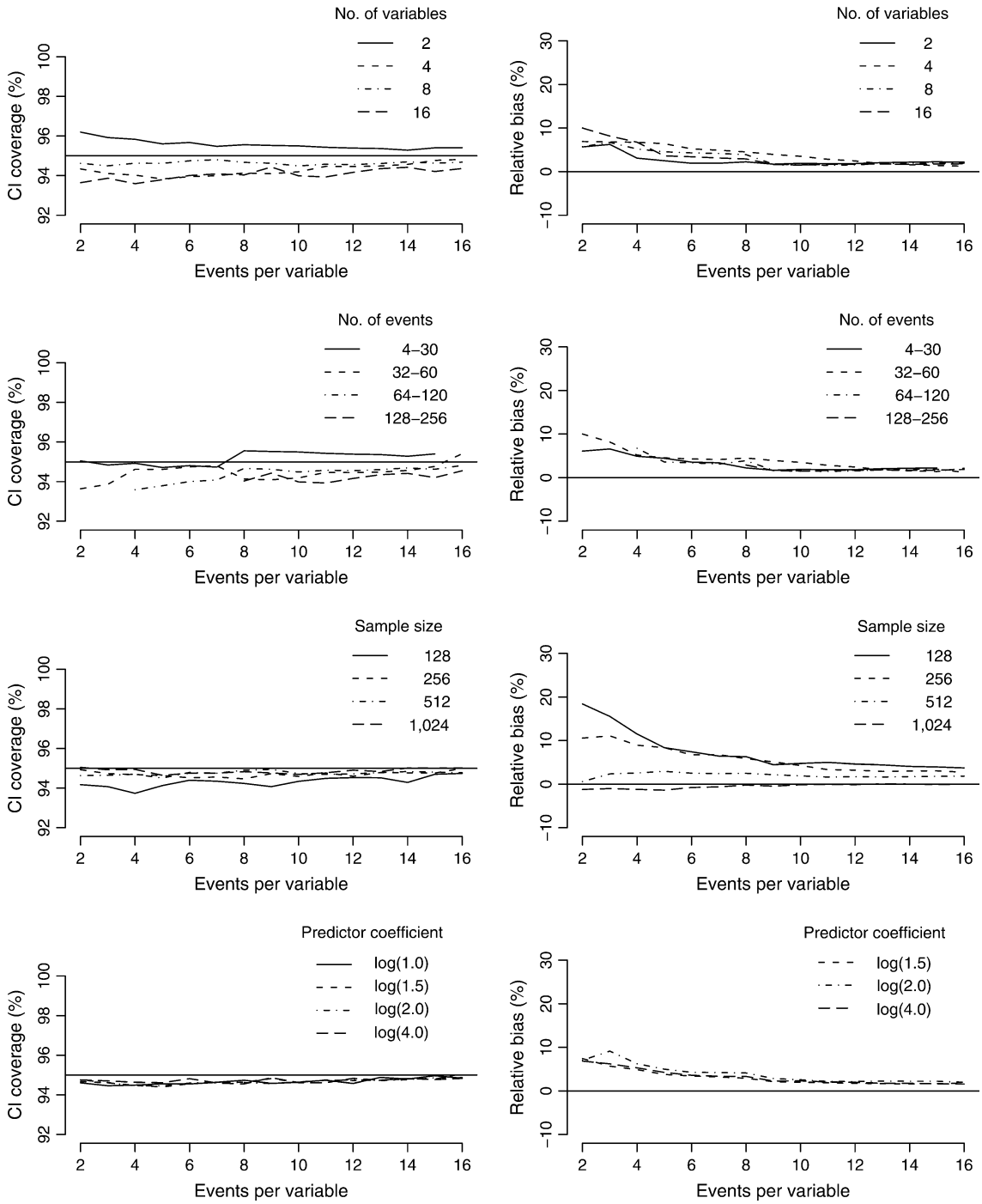

FIGURE 4. Cox model with continuous primary predictor. CI, confidence interval.

EPV from two to 16, stratified in turn by the numbers of variables, events, and observations, and then by the prevalence of the binary primary predictor or value of β_1. Averages are taken over all simulation parameters other than EPV and the stratification variable. Problem rates and worst cases are shown in tables 1 and 2, respectively, for 2–4, 5–9, and 10–16 EPV.

Logistic regression with binary primary predictor

Results are shown in figure 1. For the primary predictor, the average confidence interval coverage for β_1 was generally at or above the nominal level. The conservatism was apparent only in data sets with 30 or fewer events. Sample size did not affect confidence interval coverage. Values of

Downloaded from http://aje.oxfordjournals.org/ at Kainan University on March 1, 2015

TABLE 1. Scenarios with problematic performance*

Model	Primary predictor	Problem†	Events per variable		
			2–4	5–9	10–16
Logistic	Binary	Confidence interval coverage: <93%	5.9	1.7	1.4
		Type I error: >7%	4.4	0.9	0.7
		Relative bias: >15%	26.4	7.4	2.7
		Any of three problems	23.7	7.2	3.4
	Continuous	Confidence interval coverage: <93%	6.0	2.5	1.0
		Type I error: >7%	6.8	1.7	1.1
		Relative bias: >15%	21.0	6.1	2.8
		Any of three problems	19.3	6.9	3.2
Cox	Binary	Confidence interval coverage: <93%	8.2	5.8	3.1
		Type I error: >7%	6.4	3.4	2.3
		Relative bias: >15%	25.3	6.4	2.7
		Any of three problems	25.1	10.4	5.1
	Continuous	Confidence interval coverage: <93%	9.5	7.0	3.0
		Type I error: >7%	9.5	6.9	1.9
		Relative bias: >15%	17.2	2.0	0.0
		Any of three problems	19.8	8.6	3.0

* All measures are shown in percent.

† Confidence interval coverage and any problem are evaluated in all scenarios; type I error is evaluated only in scenarios with $\beta_1 = 0$, while relative bias is evaluated only in scenarios with $\beta_1 > 0$.

EPV were associated with confidence interval coverage when the prevalence of x_1 was 25 percent or 50 percent, but not at 10 percent. Neither the magnitude of β_1 nor the multiple correlation of x_1 with other predictors affected confidence interval coverage (results not shown). Confidence interval coverage was less than 93 percent in 1.7 percent of scenarios with 5–9 EPV, and the type I error rate was greater than 7 percent in 0.9 percent of scenarios (table 1). Minimum observed confidence interval coverage and maximum type I error rates were similar for 5–9 EPV and 10–16 EPV but considerably worse with 2–4 EPV (table 2).

We found mild relative bias in the estimate of β_1 except with 2–4 EPV; in that case, it was confined mainly to models with only two predictors and to predictors with either low (10 percent) or high (50 percent) prevalence (figure 1, right column). The upward bias with low prevalence predictors may be explained by failure to converge, which was observed in greater than 5 percent of data sets only with 2–4 EPV or 30 or fewer events (results not shown). Relative bias was greater than 15 percent in 7.4 percent of scenarios with 5–9 EPV (table 1) but generally comprised less than 10 percent of root mean squared error. Maximum bias was moderately larger with 5–9 EPV than with 10–16 EPV, but much smaller than with 2–4 EPV (table 2). Power was less than 80 percent in 80 percent of the scenarios examined, increasing as expected with the magnitude of β_1, as well as the number of events and sample size, and decreasing as the correlation of x_1 with the other predictors increased. Overall, we found problems in 7.2 percent of scenarios with 5–9 EPV (table 1), mainly in those with two predictors and 30 or fewer events.

Logistic regression with continuous primary predictor

Results are shown in figure 2. The average confidence interval coverage was within one percentage point of the nominal level in almost all circumstances, nearly constant at values of EPV greater than or equal to five, and influenced as much by the numbers of variables (first row) and events (second row) as by EPV. Coverage appeared liberal only with 16 predictors and 10 or fewer EPV. The true value of β_1 had little apparent influence, and we found no effect of the multiple correlation of x_1 with the other predictors (results not shown). Confidence interval coverage was less than 93 percent in 2.5 percent of scenarios with 5–9 EPV, and type I error was greater than 7 percent in 1.7 percent. The minimum observed confidence interval coverage and maximum type I error rates were similar for 2–4, 5–9, and 10–16 EPV.

In terms of relative bias, the influence of EPV was apparent when the number of predictor variables was small. However, sample size was considerably more influential than EPV (third row), and even with 10 or more EPV, average bias away from the null was roughly 5 percent. Relative bias

Downloaded from http://aje.oxfordjournals.org/ at Kainan University on March 1, 2015

TABLE 2. Worst observed problems*

Model	Primary predictor	Problem†	Events per variable		
			2–4	5–9	10–16
Logistic	Binary	Minimum confidence interval coverage	85.7	91.0	91.2
		Maximum type I error	14.3	7.6	7.8
		Maximum relative bias	260.1	51.3	36.8
	Continuous	Minimum confidence interval coverage	89.2	90.4	92.4
		Maximum type I error	8.6	8.6	7.6
		Maximum relative bias	65.6	40.2	35.0
Cox	Binary	Minimum confidence interval coverage	85.0	88.4	90.6
		Maximum type I error	12.9	9.0	8.6
		Maximum relative bias	240.4	51.4	40.7
	Continuous	Minimum confidence interval coverage	90.4	91.0	91.0
		Maximum type I error	8.8	8.8	8.0
		Maximum relative bias	51.8	29.5	13.9

* All measures are shown in percent.

† Confidence interval coverage is evaluated in all scenarios; type I error is evaluated only in scenarios with $\beta_1 = 0$, while relative bias is evaluated only in scenarios with $\beta_1 > 0$.

was greater than 15 percent in 6.1 percent of scenarios with 5–9 EPV, but it generally comprised no more than 10 percent of root mean squared error. Maximum bias was moderately larger with 5–9 EPV than with 10–16 EPV but also moderately smaller than with 2–4 EPV. Power was less than 80 percent in 87 percent of the scenarios examined and responded predictably to inputs. Overall, we found problems in 6.9 percent of scenarios with 5–9 EPV, mainly in those with two or 16 predictors.

Cox regression with binary primary predictor

Results are shown in figure 3. We found departures in confidence interval coverage from the nominal level in both directions. Liberal confidence intervals were observed only for models with 16 predictors. In contrast, conservatism depending on EPV was observed in models with two and four predictors. The conservatism with 10 or fewer EPV was more pronounced with larger samples. The effects of the prevalence of the predictor, as well as its multiple correlation with other predictors and the magnitude of β_1, were minor (results for the latter not shown). Confidence interval coverage was less than 93 percent in 5.8 percent of scenarios with 5–9 EPV, and type I error was greater than 7 percent in 3.4 percent. The minimum observed confidence interval coverage and maximum type I error rates were slightly worse for 5–9 EPV than for 10–16 EPV but considerably better than for 2–4 EPV.

We found some bias in β_1 with 2–4 EPV, depending on the number of predictor variables or events. Sample size had little or no apparent effect. Substantial bias away from the null was observed only with low predictor prevalence, in association with 2–4 EPV, 30 or fewer events, and resulting model convergence rates less than 95 percent. The magnitude of β_1 and the multiple correlation of x_1 with the other predictors were also influential in this range of values of EPV (results not shown). Relative bias was greater than 15 percent in 6.4 percent of scenarios with 5–9 EPV but generally comprised less than 10 percent of root mean squared error. Maximum bias was similar with 5–9 EPV and 10–16 EPV but much smaller than with 2–4 EPV. Estimated power was less than 80 percent in 74 percent of scenarios, responded predictably to inputs, and showed little dependence on sample size after the number of events was taken into account (7). Overall, we found problems in 10.4 percent of scenarios with 5–9 EPV, mainly in those with two or 16 predictors.

Cox regression with continuous primary predictor

Results are shown in figure 4. Confidence interval coverage was slightly conservative with two predictors and slightly liberal with four or more predictors. There was little or no apparent influence of EPV. The regression coefficient for x_1 and its correlation with the other predictors were similarly unimportant. Confidence interval coverage was less than 93 percent in 7.0 percent of scenarios with 5–9 EPV, and type I error was greater than 7 percent in 6.9 percent. The minimum observed confidence interval coverage and maximum type I error rates were similar for 5–9 and 10–16 EPV.

Bias away from the null in β_1 was observed with 10 or fewer EPV in this case. However, bias was less than 5 percent except with four or fewer EPV and 16 predictors or in relatively small samples of 128 or 256 observations. Bias did not strongly depend on the magnitude of β_1, nor on the

Downloaded from http://aje.oxfordjournals.org/ at Kainan University on March 1, 2015

correlation of x_1 with other predictors. Relative bias was greater than 15 percent in only 2 percent of scenarios with 5–9 EPV and generally comprised 10 percent or less of root mean squared error. Maximum bias was moderately larger with 5–9 EPV than with 10–16 EPV but also moderately smaller than with 2–4 EPV. Estimated power was less than 80 percent in 82 percent of scenarios and responded predictably to inputs. Overall, we found problems in 8.6 percent of scenarios with 5–9 EPV.

Additional simulations

To reflect the setup considered by Peduzzi et al. (2, 3) more closely, we also examined models with all binary predictors. For both the logistic and Cox models, results were very similar to those seen with continuous covariates, with confidence interval coverage, type I error rates, and relative bias for the primary predictor at most slightly degraded. In addition, for the logistic model, we also assessed bias-corrected, percentile-based bootstrap confidence intervals in selected problematic scenarios with five EPV and $n = 256$. The bootstrap confidence intervals were somewhat more conservative than the Wald confidence intervals, often with coverage greater than 95 percent.

DISCUSSION

Our simulation study shows that the rule of thumb of 10 or more EPV in logistic and Cox models is not a well-defined bright line. If we (somewhat subjectively) regard confidence interval coverage less than 93 percent, type I error greater than 7 percent, or relative bias greater than 15 percent as problematic, our results indicate that problems are fairly frequent with 2–4 EPV, uncommon with 5–9 EPV, and still observed with 10–16 EPV. Cox models appear to be slightly more susceptible than logistic. The worst instances of each problem were not severe with 5–9 EPV and usually comparable to those with 10–16 EPV.

Our evaluation focuses primarily on confidence interval coverage for β_1 and the related type I error rate of the test of H_0 ($\beta_1 = 0$), secondarily on bias in the estimate of β_1, and only indirectly on variability and power. These emphases are motivated by the fact that, in the situations we have considered, power is usually low and variability is high. However, because bias on average comprises only 10–20 percent of root mean squared error, confidence interval coverage and the type I error rate are fairly well maintained even in the presence of considerable bias. We draw three broad implications from our results.

- In this context, type II errors will be common, but misleading conclusions can usually be avoided if negative findings are interpreted in the light of confidence intervals (8) with expected coverage close to the nominal level. Our results show that these conditions usually hold with five or more EPV.
- Mildly conservative confidence intervals and type I error rates were the dominant pattern even when parameter estimates were biased away from the null. This implies that, when a statistically significant association is found in a model with 5–9 EPV, only a minor degree of extra caution is warranted, in particular for plausible and highly significant associations hypothesized a priori.
- If even the low risk of problems seen with 5–9 EPV is unacceptable, modern resampling tools can be used to validate the model-based inferences. For example, the bootstrap can be used to assess bias and frequency of nonconvergence and to derive bias-corrected confidence intervals.

Our results suggest other contexts in which extra caution in interpretation is warranted. For example, the confidence interval coverage was eroded in larger models, especially at low EPV. Bias away from the null was also exacerbated with continuous primary predictors by smaller sample sizes and with binary primary predictors by low predictor prevalence. The latter stems from the fact that, when no events are observed in the small set of "exposed" observations, the model does not converge.

Our simulation study, while large, has limitations. In particular, our graphical summaries averaging over parameters other than EPV and a single stratification variable may obscure some circumstances in which confidence interval coverage or bias is considerably worse than the average. However, our tabulation shows that such problems are uncommon, usually not severe, and are also observed with 10 or more EPV.

Bigger samples and more events are almost always preferable. However, situations commonly arise where confounding cannot be persuasively addressed without violating the rule of thumb we have studied. In that case, we agree with Peduzzi et al. (2) that results should be interpreted with caution and, in addition, compared with those from models from which weaker predictors have been excluded. However, systematic discounting of results, in particular statistically significant associations, from any model with 5–9 EPV does not appear to be justified.

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES

1. Concato J, Peduzzi P, Holfold TR, et al. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. J Clin Epidemiol 1995;48:1495–501.
2. Peduzzi P, Concato J, Feinstein AR, et al. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;48:1503–10.
3. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373–9.
4. Peduzzi P, Detre K, Gage A. Veterans Administration Cooperative Study of medical versus surgical treatment for

Downloaded from http://aje.oxfordjournals.org/ at Kainan University on March 1, 2015

stable angina—progress report. Section 2. Design and baseline characteristics. Prog Cardiovasc Dis 1985;28: 219–28.
5. Harrell FE, Lee KL, Mark DB. Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15:361–87.
6. Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health 1989;79:340–9.
7. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. Biometrics 1983;39:499–503.
8. Hoenig JM, Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. Am Stat 2001;55:19–24.

Downloaded from http://aje.oxfordjournals.org/ at Kainan University on March 1, 2015