

THE RELATIVE IMPORTANCE OF PROGNOSTIC FACTORS IN STUDIES OF SURVIVAL*

M. SCHEMPER

*Section of Clinical Biometry, Department of Medical Computer Sciences, Vienna University, Garnisongasse 13,
A-1090 Vienna, Austria*

SUMMARY

The relative importance of prognostic factors in regression can be measured either by standardized regression coefficients or by percentages of explained variation in a dependent variable. One advantage of using explained variation is the direct comparability of qualitative prognostic factors with others, or of groups of prognostic factors. The description of relative importance can be accomplished within marginal or partial effects analyses. It is demonstrated that it is possible not only to provide a descriptive ranking of prognostic factors according to their statistically determined importance, but also to make inferences concerning their relative importance, employing bootstrap techniques and procedures for multiple comparisons. The methods presented, which are new in the context of Cox regression, are exemplified by analyses of studies of lung cancer and breast cancer.

1. INTRODUCTION

Standards for the analysis of prognostic factors were set long ago (for example by Armitage and Gehan¹ and Gehan and Walker²). These standards comprise marginal and partial analyses by Cox³ and logistic⁴ regression, checking for interaction and for non-linear or time-dependent effects for each factor, and presentation of results in terms of estimated relative risks, corresponding confidence intervals and *P*-values. Occasionally however it is necessary to comment on the relative importance of factors or on the relative weight of groups of factors within a model. Relative importance has been addressed by very few papers but has been well reviewed recently by Healy.⁵

In this paper procedures to supplement standard analyses of prognostic factors will be discussed. These aim to quantify knowledge about the outcome of a disease and the relative weight of individual factors in determining this outcome.

Measures used to describe the relative importance of prognostic factors in regression models are explained variation and standardized regression coefficients. Both lead to similar results for continuous or dichotomous factors. However if qualitative and quantitative factors or groups of factors are compared then explained variation is better, as the effect of any factor or group of factors can be characterized by a single number; this is not possible with regression coefficients. Only explained variation will be considered here.

* Presented at the Society for Clinical Trials International Society for Clinical Biostatistics Joint Meeting, Brussels, Belgium, July 1991.

In linear regression models squared multiple correlation, R^2 , is a popular measure of the proportion of variation of a dependent variable explained (PVE) by prognostic factors. For Cox³ models of survival data, useful analogues to R^2 have become available only recently.

2. MEASURES OF EXPLAINED VARIATION IN COX MODELS

The first useful PVE measure, V_2 , was derived by the author⁶ in 1990. It measures the relative discrepancy of individual survival from Cox model survival functions compared with the discrepancy from Kaplan–Meier⁷ estimates. The measure V_2 is intuitively appealing, its validity is unaffected by violations of model assumptions, and a corresponding SAS Fortran program is available from the author. However a computationally quicker alternative to V_2 , R_M^2 , is available which is of special interest if computationally intensive methods are applied to estimates of explained variation.

The expression $R_M^2 = 1 - (L_R/L_U)^{2/n}$, where L_R and L_U are the restricted and unrestricted maximum likelihoods and n denotes total sample size, has been attributed to Maddala.⁹ The restricted likelihood L_R , maximized subject to all effect parameters being zero, is often the starting value in iterative maximum likelihood procedures.

For the standard linear regression model Magee³ demonstrated the equivalence of multiple R^2 , R_M^2 and $R_{LR}^2 = 1 - \exp(-LR/n)$, where $LR = 2 \log(L_U/L_R)$. While these statistics are all available for multiple linear regression, only R_{LR}^2 and R_M^2 can be used conveniently with more general regression models. The following guidelines for using R_M^2 in Cox model analyses have been made.¹⁰ The total sample size rather than the number of uncensored failure times should be substituted for n in the above definitions of R_M^2 and R_{LR}^2 . Contrary to V_2 the validity of R_M^2 does depend on the proportionality assumption of the Cox model, but V_2 and R_M^2 lead to identical results if proportionality of hazards holds approximately. Therefore V_2 is preferred to R_M^2 only with marked non-proportionality of hazards. In the following sections only R_M^2 will be used.

R_M^2 can be calculated for a full model as well as for each prognostic factor, whether quantitative or qualitative. It is sensible to describe the PVE by individual factors both in a simple, unadjusted, marginal and in an adjusted, partial sense. Results for marginal PVE of factors are obtained by separate regressions for each prognostic factor, while partial PVE is calculated as the difference of R_M^2 for the full model and R_M^2 for a model with the factor of interest excluded.

The prognostic importance of each factor can now be quantified by marginal or partial R_M^2 s, which also determine the order of the factors' importance.

3. INFERENCE FOR A RANKING BY IMPORTANCE

In this section we present methods to clarify whether or not observed differences in marginal or partial explained variation between prognostic factors can be attributed to chance. For this purpose the construction of covariance matrices for marginal or partial R_M^2 by bootstrap methods^{11,12} and aspects of multiple testing will be discussed.

The bootstrap estimate of the covariance matrix can be thought of as a non-parametric analogue of maximum likelihood. The steps to arrive at this covariance matrix for the R_M^2 s (of dimension (k, k) for k prognostic factors) are sketched briefly as follows:

1. Take a random sample of size n with replacement from the given study sample of size n to be analysed.
2. For this random sample calculate the chosen measure of importance M_j , $1 \leq j \leq k$ (marginal or partial R_M^2) for each of the k prognostic factors after fitting separate Cox models. It is

recommended to take $M_j = \arcsin \sqrt{R_M^2}$, the distribution of which is closer to normal than that of the untransformed percentage of explained variation R_M^2 .

- Repeat steps 1 and 2 N times (for example $N = 100$; cf. Efron¹²) and from these N repetitions calculate variances and covariances ($\text{cov}_{jj'}$, $1 \leq j, j' \leq k$) of the M_j for all k prognostic factors. For the comparison of factor j with factor j' ,

$$(M_j - M_{j'}) / \sqrt{(\text{cov}_{jj} + \text{cov}_{j'j'} - 2\text{cov}_{jj'})}$$

is compared with the quantiles of the standard normal distribution.

Alternatively and computationally quicker, instead of bootstrapping to obtain the covariance matrix of the M_j ($1 \leq j \leq k$), the variances and covariances for the differences $M_j - M_{j'}$ ($1 \leq j \leq j' \leq k$) can be obtained by the bootstrap directly. This covariance matrix is denoted by $\text{cov}(D)_{ll'}$ with $1 \leq l, l' \leq k(k-1)/2$, the upper limit giving the number of pairwise comparisons for k prognostic factors.

Furthermore, we use mean R_M^2 s from the bootstrap replications and report these as estimates of PVE rather than the observed R_M^2 s.

The methods described so far permit the testing of hypotheses specified in advance and without considering multiplicity. An example is presented in Section 4, where it is of interest to find out which of two prognostic markers should be obtained routinely to improve prediction of survival. See also an example discussed by Cohen.¹³

In a study of prognostic factors however all would be contrasted, resulting in similar violations of the global significance level as uncorrected pairwise comparisons of means in a k -sample setting. To find out which prognostic factors differ in importance at a prespecified significance level α , methods for multiple comparisons or simultaneous inference are required.

The easiest method is the Bonferroni¹⁴ correction where the observed error probabilities of all comparisons are multiplied by $k(k-1)/2$, the number of comparisons, and only differences are assumed significant at prespecified α (for example $\alpha = 0.05$) where these products remain below α . More efficient than the simple Bonferroni correction are related techniques by Holm¹⁵ and Shaffer,¹⁶ which are applicable in this context. If multiple comparisons are tailored to a specific correlation matrix then integration of a multivariate normal distribution is needed to obtain the critical limit for each comparison under a global level α (cf. Reference 14, p. 365).

If sample sizes are reasonably large, the number of factors is kept to a reasonable limit and high correlations between factors are avoided, then it is reasonable to apply the Bonferroni correction and avoid numerically delicate approaches for which only limited experience is available. High correlations can be avoided by selecting a single factor from those which are highly correlated.

4. EXAMPLES

4.1. Analysis of prognostic factors in a study of lung cancer

The Veteran's Administration lung cancer data published in Kalbfleisch and Prentice¹⁷ provide a typical sample for an analysis of prognostic factors. They consist of 137 patients for whom 128 uncensored and 9 censored survival times are recorded. Four prognostic factors were considered, two continuous (age and Karnofsky performance index), one dichotomous (treatment) and one qualitative (histology, four levels). The results of marginal and partial effect analyses are given in Table I. The strength of the factors is described by conventional estimates of the relative risk from Cox analyses, accompanied by confidence intervals and P -values. However it is difficult to compare the estimates because of different scales of measurement and different types of factors.

Table I. VA study of lung cancer¹⁷

Factors	Relative risk (CI)	P	PVE (total = 38.9%)
<i>Marginal effects analysis</i>			
Treatment	1.02 (0.71-1.45)	0.92	0.0%
Age	1.01 (0.99-1.03)	0.43	1.3%
Histology	2.7 3.1 1.3 (1.65-4.45 1.76-5.56 0.73-2.17)	< 0.001	17.4%
Karnofsky index	0.97 (0.96-0.98)	< 0.001	26.7%
<i>Partial effects analysis</i>			
Treatment	1.3 (0.90-2.01)	0.15	1.3%
Age	0.99 (0.97-1.01)	0.34	0.8%
Histology	2.3 3.2 1.5 (1.38-3.98 1.81-5.77 0.86-2.60)	< 0.001	10.4%
Karnofsky index	0.97 (0.96-0.98)	< 0.001	19.2%

PVE = proportion of variation explained.
CI = 95% confidence interval.

With treatment the relative risk estimate refers to the differential effect of treatment, while with age the change in risk between two successive years of age is expressed. Furthermore the strength of the histology factor can only be summarized by three separate estimates.

If the prognostic importance of factors is to be compared then comparisons in terms of PVE (R_M^2) are most suitable.

Applying the simple Bonferroni adjustment, all differences in the importance of the factors are confirmed for the marginal analysis at a significance level $\alpha = 0.01$, except for treatment versus age and histology versus Karnofsky index. In the corresponding partial analysis only the comparisons of Karnofsky index with treatment and with age are confirmed. It should be noted that the PVE for the full model is 38.9 per cent, while the sums of marginal and partial PVEs lie above and below this value respectively.

4.2. Relative importance of two prognostic markers in breast cancer

In a sample of 175 patients who received primary treatment for breast cancer and were recruited for an oncology trial between 1977 and 1982 at the 1st Department of Surgery, Vienna University, survival times (78 uncensored) and standard prognostic factors were recorded. Average age of these patients was 55.6 years, estimated median follow-up time 12.4 years. Two dominant prognostic factors, lymph nodes (positive versus negative) and tumour grading (3 versus 1 or 2), explained 20 per cent of variation in survival. A medical investigator obtained values of two prognostic markers from flow cytometry, aneuploidy fraction (PLO) and DNA index (DNI), for these patients.¹³ He was interested in whether PLO and DNI are of independent prognostic importance, in addition to lymph nodes and grading, and if either PLO or DNI is significantly more important than the other. In other words, if only one of the markers is used routinely, can this choice be based on confirmed superiority?

It turned out in a standard Cox analysis (see Table II) that the additional prognostic information carried by PLO is confirmed ($p = 0.0004$) while that of DNI is not ($p = 0.095$). The respective PVE increases from 20 to 25 per cent for PLO and to 22 per cent for DNI. However there is only

Table II. Two prognostic markers, DNI and PLO, in breast cancer

Factors	Relative risk (CI)	P	PVE
Grading	2.08 (1.28-3.38)	0.010	20%
Lymph nodes	4.10 (2.46-6.83)	< 0.001	
DNI	1.53 (0.93-2.51)	0.095	8%
PLO	2.34 (1.47-3.73)	< 0.001	5%

Results for grading and lymph nodes are mutually adjusted. Results for DNI and PLO respectively are adjusted for grading and lymph nodes.

an indication ($p = 0.092$) that PLO is preferable over DNI, and based on the given sample, the observed differences in prognostic importance might still be due to chance. The latter answer was obtained by bootstrapping PVE for a partial analysis of DNI and of PLO - partial with respect to the presence of lymph nodes and grading in both models.

5. CONCLUDING REMARKS

The preceding examples demonstrate typical and straightforward applications. The general principles of conventional analysis of prognostic factors still apply when effects are summarized by PVE instead of relative risk. In many applications the mechanisms influencing outcome may be too complex to be summarized by a model containing main effects only. In such cases it is possible to estimate PVE for interaction effects, non-linear effects or time-dependent effects. If such effects become dominant, for example with qualitative or strong quantitative interactions, then, as with conventional analysis, evaluation of PVE by subgroups may be indicated. Occasionally, in particular with highly correlated interacting factors, it makes sense to evaluate PVE for groups of such factors or for an aggregated factor, constructed from others.

Analysis of the relative importance of prognostic factors is of interest in many biomedical research contexts and PVE currently seems to be the most suitable measure. Various criticisms¹⁹⁻²³ of multiple R^2 and of related definitions of PVE focus on its use as a measure of goodness-of-fit or its use under very special conditions, but these criticisms do not apply to PVE as a measure of explained risk (see also Korn and Simon²⁴) and thus are not relevant to applications in prognostic factor studies.

On the other hand the use of PVE measures reveals certain properties of a data set not adequately expressible in terms of estimates of regression parameters or P -values. With these techniques the predictive power of prognostic factors has often been overestimated. As pointed out by Korn and Simon,²⁵ this overestimation may discourage search for better prognostic factors, it may encourage the use of new diagnostic assessments of low practical value, and it may give too much credence to non-randomized studies of treatment, their analysis being based on adjustments by prognostic factors of little predictive power.

In this paper emphasis has been on applications to Cox models as the use of PVE measures is still uncommon in this area. However the ideas presented apply to other regression models as well.

The procedures described are embodied in an SAS macro called RELIMP (RELative IMPortance) which is available on request.

REFERENCES

1. Armitage, P. and Gehan, E. A. 'Statistical methods for the identification and use of prognostic factors', *International Journal of Cancer*, 13, 16-36 (1974).

2. Gehan, E. A. and Waiker, M. D. 'Prognostic factors for patients with brain tumors'. *National Cancer Institute Monographs*, **46**, 189-195 (1978).
3. Cox, D. R. 'Regression models and life tables'. *Journal of the Royal Statistical Society, Series B*, **34**, 187-220 (1972).
4. Cox, D. R. *The Analysis of Binary Data*, Methuen, London, 1970.
5. Healy, M. J. R. 'Measuring importance'. *Statistics in Medicine*, **9**, 633-637 (1990).
6. Schemper, M. 'The explained variation in proportional hazards regression'. *Biometrika*, **77**, 216-218 (1990).
7. Kaplan, E. L. and Meier, P. M. 'Nonparametric estimation from incomplete observations'. *Journal of the American Statistical Association*, **53**, 457-481 (1958).
8. Maddala, G. S. *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, 1983.
9. Magee, L. ' R^2 measures based on Wald and likelihood ratio joint significance tests'. *The American Statistician*, **44**, 250-253 (1990).
10. Schemper, M. 'Further results on the explained variation in proportional hazards regression'. *Biometrika*, **79**, 202-204 (1992).
11. Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM, Philadelphia, 1982.
12. Efron, B. 'Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods'. *Biometrika*, **68**, 589-599 (1981).
13. Cohen A. 'Comparison of correlated correlations'. *Statistics in Medicine*, **8**, 1485-1495 (1989).
14. Hochberg, Y. and Tamhane, A. C. *Multiple Comparison Procedures*, Wiley, New York, 1987.
15. Holm, S. 'A simple sequentially rejective multiple test procedure'. *Scandinavian Journal of Statistics*, **6**, 65-70 (1979).
16. Shaffer, J. P. 'Modified sequentially rejective multiple test procedures'. *Journal of the American Statistical Association*, **81**, 826-831 (1986).
17. Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1980.
18. Gnant, M. and 14 coauthors. 'Ploidy index: a new prognostic factor in operable human breast cancer - 10 years results' (abstract), 5th EORTC Breast Cancer Meeting, Leuven, Belgium, 1991.
19. Kvalseth, T. O. 'Cautionary note about R^2 '. *The American Statistician*, **39**, 279-285 (1985).
20. Willett, J. B. and Singer, J. D. 'Another cautionary note about R^2 : its use in weighted least-squares regression analysis'. *The American Statistician*, **42**, 236-238 (1988).
21. Scott, A. and Wild, C. 'Transformations and R^2 '. *The American Statistician*, **45**, 127-129 (1991).
22. Healy, M. J. R. 'The use of R^2 as a measure of goodness of fit'. *Journal of the Royal Statistical Society, Series A*, **147**, 608-609 (1984).
23. Helland, I. S. 'On the interpretation and use of R^2 in regression analysis'. *Biometrics*, **43**, 61-69 (1987).
24. Korn, E. L. and Simon, R. 'Explained residual variation, explained risk, and goodness of fit'. *The American Statistician*, **45**, 201-206 (1991).
25. Korn, E. L. and Simon, R. 'Measures of explained variation for survival data'. *Statistics in Medicine*, **9**, 487-503 (1990).