STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:479–489
Published online 8 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.1946

Explained randomness in proportional hazards models

John O'Quigley[1,‡], Ronghui Xu[2,*,†] and Janez Stare[3,§]

[1]*Institut Curie, Paris*
[2]*Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, U.S.A.*
[3]*Department of Biomedical Informatics, School of Medicine, University of Ljubljana, Slovenia*

SUMMARY

A coefficient of explained randomness, analogous to explained variation but for non-linear models, was presented by Kent. The construct hinges upon the notion of Kullback–Leibler information gain. Kent and O'Quigley developed these ideas, obtaining simple, multiple and partial coefficients for the situation of proportional hazards regression. Their approach was based upon the idea of transforming a general proportional hazards model to a specific one of Weibull form. Xu and O'Quigley developed a more direct approach, more in harmony with the semi-parametric nature of the proportional hazards model thereby simplifying inference and allowing, for instance, the use of time dependent covariates. A potential drawback to the coefficient of Xu and O'Quigley is its interpretation as explained randomness in the covariate given time. An investigator might feel that the interpretation of the Kent and O'Quigley coefficient, as a proportion of explained randomness of time given the covariate, is preferable. One purpose of this note is to indicate that, under an independent censoring assumption, the two population coefficients coincide. Thus the simpler inferential setting for Xu and O'Quigley can also be applied to the coefficient of Kent and O'Quigley. Our second purpose is to point out that a sample-based coefficient in common use in the SAS statistical package can be interpreted as an estimate of explained randomness when there is no censoring. When there is censoring the SAS coefficient would not seem satisfactory in that its population counterpart depends on an independent censoring mechanism. However there is a quick fix and we argue in favour of its use. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: correlation; explained randomness; information gain; proportional hazards

1. BACKGROUND AND MOTIVATION

The main random variable of interest is time T, defined on the interval $(0, \infty)$. Our interest focuses on T and how it relates to a p vector of explanatory variables Z. The data will

*Correspondence to: Ronghui Xu, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, U.S.A.
†E-mail: rxu@jimmy.harvard.edu
‡E-mail: joq@biomath.jussieu.fr
§E-mail: janez.stare@mf.uni-lj.si

Received January 2003
Accepted May 2004

be of the form of n replicates (T_i, C_i, Z_i) of the vector (T, C, Z) in which C is a censoring variable, independent of T given Z (although we sometimes need a stronger assumption of marginal independence). Time dependency is indicated via $Z = Z(t)$. In practice, for each i, we observe $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leqslant C_i)$, where $I(\cdot)$ is the indicator function. Let $Y_i(t)$ denote whether the subject i is at risk ($Y_i(t) = 1$) or not ($Y_i(t) = 0$) at time t. Let $N_i(t) = I\{T_i \leqslant t, T_i \leqslant C_i\}$ and $\bar{N}(t) = \sum_1^n N_i(t)$. The total number of observed failures is denoted by $k = \bar{N}(\infty)$. The left continuous version of the Kaplan–Meier estimator of survival is denoted $\hat{S}(t)$ and the Kaplan–Meier estimate of the distribution function of T by $\hat{F}(t) = 1 - \hat{S}(t)$. The proportional hazards regression model [1] specifies that the hazard function can be written as;

$$\lambda\{t \mid Z(t)\} = \lambda_0(t) \exp\{\beta' Z(t)\} \tag{1}$$

where $\lambda_0(t)$ is a fixed but unknown 'baseline' hazard function, and β is a $p \times 1$ regression parameter to be estimated. The model-based quantities $\pi_j(t; \beta)$, defined by;

$$\pi_j(t; \beta) = \frac{Y_j(t) \exp(\beta' Z_j(t))}{\sum_{l=1}^n Y_l(t) \exp(\beta' Z_l(t))} \tag{2}$$

are simply the conditional probabilities of choosing individual j, given all the individuals at risk at time t and that one individual is to be selected to fail. The product of the π's over the observed failure times is the so-called partial likelihood [1, 2], which, as a function of β, is maximized at the partial likelihood estimate $\hat{\beta}$. We have, for our purposes, a stronger result; the $\pi_j(t; \beta)$'s are not only conditional probabilities as described above, the $\pi_j(t; \hat{\beta})$'s provide, in addition, consistent estimates of the conditional distribution of Z given T under model (1). This is summarized in a theorem of Xu and O'Quigley [3].

Theorem 1
Under model (1), the conditional distribution function of Z given T is consistently estimated by

$$\hat{P}(Z(t) \leqslant z \mid T = t) = \sum \pi_j(t; \hat{\beta}) I(Z_j(t) \leqslant z)$$

In the above '$\leqslant$' denotes component-wise less than or equal to. Before recalling, below, the concept of information gain and explained randomness, we proceed, somewhat out of order, by indicating the properties that make it attractive. A coefficient of explained randomness of T given Z, ρ^2, developed by Kent and O'Quigley [4] has the following properties: (1) When a covariate Z is unrelated to survival, and the corresponding regression coefficient is zero, then $\rho^2 = 0$. (2) When $\beta \neq 0$ then $0 < \rho^2 < 1$. Following Kent [5], non-zero values of ρ^2 are ordered in terms of the proportion of explained randomness of T given Z. (3) ρ^2 is invariant under linear transformations of Z and under monotone increasing transformations of T. (4) For one-dimensional β and fixed covariate distribution, ρ^2 is an increasing function of $|\beta|$. (5) ρ^2 remains invariant under all independent censoring mechanisms. (6) The estimate $\hat{\rho}^2$ proposed by Kent and O'Quigley is consistent for ρ^2 and is asymptotically normal.

The coefficient presented by Xu and O'Quigley [3], written here as ρ^2_{XOQ}, used the same basic ideas but, exploiting the above theorem and looking at the distribution of Z at each time $T = t$, the construction could be carried out using routine quantities calculated during a standard proportional hazards analysis. Inference was also greatly simplified. In addition the

presence of time-dependent covariates presents no difficulties for the coefficient ρ^2_{XOQ} and its sample-based estimate whereas that of Kent and O'Quigley [4] is not defined in this case. All of the above properties of ρ^2 also hold for ρ^2_{XOQ} [3].

In the next section, we review the concept of information gain and explained randomness, and the measures ρ^2 and ρ^2_{XOQ}. We then show that, under broad conditions, ρ^2 and ρ^2_{XOQ} coincide. Thus the more readily available quantities, used by Xu and O'Quigley [3], obtained from standard proportional hazards regression, can also be applied to ρ^2. Furthermore, a widely available *ad hoc* correlation measure for the proportional hazards model, suggested by a SAS survival analysis book, can find a justification, at least for uncensored data, as an alternative estimate to ρ^2. In the presence of censoring the SAS coefficient is no longer consistent, converging to zero, regardless of population effects, as the percentage of censoring increases. This problem is very easily fixed however and we also describe this in the following section.

2. INFORMATION GAIN AND EXPLAINED RANDOMNESS

Rather than work with hazards we rewrite model (1) in terms of the density so that

$$f(t\,|\,z;\beta)=\lambda_0(t)\exp\left\{\beta' z-\mathrm{e}^{\beta' z}\int_0^t\lambda_0(u)\,\mathrm{d}u\right\} \tag{3}$$

We can consider the true population effect to be quantified by the parameter β. For our specific application we will need to think in terms of different model parameterizations, indicated by the parameter θ. Mostly, the only values of θ of interest to us are those corresponding to the true model, $\theta=\beta$, our estimated model $\theta=\hat{\beta}$ and the null model of no association between T and Z, $\theta=0$. A population measure of the strength of association, or the distance between the two models indexed by $\theta=0$ and β, can be provided by twice the Kullback–Leibler information gain. This measure of information gain $\Gamma(\beta)$ expresses the expectation of the difference between the information corresponding to $\theta=\beta$ and 0. All expectations are evaluated with respect to true model indexed by β. Specifically we define; $\Gamma(\beta)=2\{I(\beta)-I(0)\}$ where

$$I(\theta)=E\{\log f(T\,|\,Z;\theta)\}=\int_{\mathscr{Z}}\int_{\mathscr{T}}\log\{f(t\,|\,z;\theta)\}f(t\,|\,z;\beta)\,\mathrm{d}t\,\mathrm{d}G(z) \tag{4}$$

In the above expression the domains of definition of T and Z are denoted by $\mathscr{T}$ and $\mathscr{Z}$, respectively, and $G(z)$ is the marginal distribution function of Z. Compared to the regression coefficient β, an information gain measure has the advantage of not depending on scale and the simple transformation

$$\rho^2=1-\exp\{-\Gamma(\beta)\} \tag{5}$$

produces a coefficient with useful interpretability properties [5]. This follows from Kent's [5] definition of the total randomness of T, as a monotonic transformation of the entropy of T which is $D(T)=\exp\{-2I(0)\}$. The residual randomness of T given Z is then $D(T\,|\,Z)=\exp\{-2I(\beta)\}$. Then $\rho^2=\{D(T)-D(T|Z)\}/D(T)$ can be interpreted as the 'proportion of the randomness' explained by the regression. For normal models and maximum likelihood or least squares estimation, $\hat{\rho}^2$ is the usual coefficient of correlation squared, equivalently the

proportion of explained variation, when, instead of working with $f(t \mid z; \beta)\, \mathrm{d}t\, \mathrm{d}G(z)$ we use the observed empirical distribution of (T, Z).

Partial measures of explained randomness in T given Z_2, after having already accounted for Z_1 can be defined directly [4]. However, it seems more satisfactory to define the partial coefficients as ratios of multiple coefficients of different orders. Thus, we work with $\rho^2(Z_1, \ldots, Z_p)$ and $\rho^2(Z_1, \ldots, Z_q)$ $(q < p)$, the multiple coefficients for models including covariates Z_1–Z_p and covariates Z_1–Z_q, respectively. We are then able to define the partial coefficient $\rho^2(Z_{q+1}, \ldots, Z_p \mid Z_1, \ldots, Z_q)$, the explained randomness provided by Z_{q+1}–Z_p after having accounted for the effects of Z_1–Z_q. The expression;

$$1 - \rho^2(Z_1, \ldots, Z_p) = [1 - \rho^2(Z_1, \ldots, Z_q)][1 - \rho^2(Z_{q+1}, \ldots, Z_p \mid Z_1, \ldots, Z_q)] \tag{6}$$

motivated by an analogous formula for the multivariate normal model, makes intuitive sense in that the value of the partial coefficient increases as the difference between the multiple coefficients increases, and takes the value zero should this difference be zero.

The following subsections consider the specific case of explained randomness for proportional hazards models that have previously been proposed.

2.1. *Kent and O'Quigley measure of explained randomness*

The survival analysis context adds a complication due to the presence of censoring. When there is no censoring then a standard estimate of information gain will be provided by n^{-1} times the usual likelihood ratio statistic [5]. Explained randomness is then estimated by using equation (5). This works for any parametric model $f(t \mid z)$ we may be interested in and the exercise is straightforward as long as we can write down $f(t \mid z)$ explicitly. In order to address the issue of censoring, Kent and O'Quigley [4] appeal to an alternative estimate using the concept of fitted information gain [6]. For fitted information gain, $I(\theta)$ is estimated by

$$\hat{I}(\theta) = n^{-1} \sum_{i=1}^{n} \int_{\mathscr{T}} \log\{f(t \mid Z_i; \theta)\} f(t \mid Z_i; \hat{\beta})\, \mathrm{d}t \tag{7}$$

with $\theta = 0$ and $\hat{\beta}$, where $\hat{\beta}$ is a consistent estimate of β. In (7), the marginal distribution of Z has been replaced by its empirical estimate. Kent [6] carried out a study on the relative merits of the two estimates of information gain. It turns out that there is little to choose between them in practice apart from the fact that the variable t enters into the expression as a dummy variable. This is important since the actual values, some of which we might not have been able to observe due to right censoring, do not have a direct effect on the calculation. The censored values have an indirect effect via their impact upon the estimation of β. We can thus readily apply the concept of fitted information gain to right censored data. This observation motivated the development of the measure of dependence for censored survival data [4] in which the conditional distribution of T can be taken to be of the Weibull form, following an unspecified monotonic transformation on T. The resulting integrals can be worked out explicitly [4, p. 528]. Thus expression (7) can be evaluated leading to an estimate of ρ^2. The calculations are somewhat involved numerically although a computer programme has been recently made available [7].

2.2. Xu and O'Quigley's measure of explained randomness ρ^2_{XOQ}

By including an unspecified baseline hazard in model (1), dependency is more readily reflected in the conditional distribution of Z given T rather than the other way around. This becomes clear when considering the usual estimating equations based on partial likelihood [8]. Indeed, for applications of the proportional hazards model in epidemiology, interest focuses directly on the conditional distribution of Z given $T=t$, where, mostly, t corresponds to age. As a consequence, Xu and O'Quigley [3] argued that, instead of (4), we might consider an alternative definition;

$$\tilde{I}(\theta)=\int_{\mathscr{T}}\int_{\mathscr{Z}}\log\{g(z\mid t;\theta)\}g(z\mid t;\beta)\,\mathrm{d}z\,\mathrm{d}F(t) \tag{8}$$

in which $F(t)$ is the marginal distribution function of T, and $g(z\mid t;\cdot)$ is the conditional density or conditional probability function of Z given T. As above we take $\tilde{\Gamma}(\beta)=2\{\tilde{I}(\beta)-\tilde{I}(0)\}$ and $\rho^2_{\mathrm{XOQ}}=1-\exp\{-\tilde{\Gamma}(\beta)\}$. We proceed in a semi-parametric way by working with a consistent estimate of F in the presence of censoring. Then an estimate of the information gain can be obtained through a semi-parametric estimator of the conditional distribution of Z given T.

Under independent censoring we can estimate the conditional distribution of Z given T by $\{\pi_j(t;\hat{\beta})\}_j$ according to Theorem 1, and the marginal distribution of T by the Kaplan–Meier estimate. The estimate can be extended to conditional independent censoring, in a way similar to Xu and O'Quigley [9]. We let $W(X_i)=\hat{F}(X_i+)-\hat{F}(X_i)$ be the jump of the Kaplan–Meier curve at an observation time X_i; this is equivalent to weighting by the inverse probability of not being censored. Then

$$\Gamma(\beta)=2\int_{\mathscr{T}}\int_{\mathscr{Z}}\log\left\{\frac{g(z\mid t;\beta)}{g(z\mid t;0)}\right\}g(z\mid t;\beta)\,\mathrm{d}z\,\mathrm{d}F(t) \tag{9}$$

can be consistently estimated using Theorem 1 and routine quantities available during a standard proportional hazards analysis. This results in the simple expression;

$$\hat{\Gamma}(\hat{\beta})=2\sum_{i=1}^{n}W(X_i)\sum_{j=1}^{n}\pi_j(X_i;\hat{\beta})\log\left\{\frac{\pi_j(X_i;\hat{\beta})}{\pi_j(X_i;0)}\right\} \tag{10}$$

The above quantity is divided by $\sum_1^n W(X_i)$ when this total probability is less than one [3]. Further, algebraic simplification was also carried out by Xu and O'Quigley [3] but is not needed for the points we wish to make here. From the above it is clear that we can achieve great simplification when compared with the approach taken to the problem of Kent and O'Quigley [4]. The results of Xu and O'Quigley [3] indicate that this second measure may be anticipated to be fairly close to the first for time-independent covariates. More light is shed on this below.

3. APPROXIMATION OF ρ^2 BY ρ^2_{XOQ}

Under certain conditions $\rho^2_{\mathrm{XOQ}}=\rho^2$. These conditions are satisfied by a bivariate normal model (itself not in the proportional hazards class) and we only expect them to be approximately met

in practice. However, on the basis of this, we anticipate ρ^2 and ρ^2_{XOQ}, as well as their sample-based estimates to be generally close. We first define $W = \beta' Z$ to be the usual prognostic index, then; we have;

Theorem 2
If, for each t, $E\{Z(\partial f(t|Z)/\partial W) = 0\}$ then, $\rho^2_{\mathrm{XOQ}} = \rho^2$.

Proof
The result follows if $I(\theta) = \tilde{I}(\theta) + K$ where K is a constant. Letting $\mathrm{d}H(z,t;\beta) = g(z \mid t;\beta)\,\mathrm{d}z\mathrm{d}F(t) = f(t \mid z;\beta)\,\mathrm{d}t\,\mathrm{d}G(z)$ then, from equation (8) we can write

$$\begin{aligned}\tilde{I}(\theta) &= \int_{\mathscr{T}}\int_{\mathscr{Z}} \log\{g(z \mid t;\theta)\}\,\mathrm{d}H(z,t;\beta) \\ &= \int_{\mathscr{T}}\int_{\mathscr{Z}} \{\log f(t \mid z;\theta) + \log g(z) - \log f(t;\theta)\}\,\mathrm{d}H(z,t;\beta) \\ &= I(\theta) + K_1 - B(\theta) \end{aligned} \tag{11}$$

where $B(\theta) = \int_{\mathscr{T}}\int_{\mathscr{Z}} \log f(t;\theta)\,\mathrm{d}H(z,t;\beta)$ and $K_1 = \int_{\mathscr{T}}\int_{\mathscr{Z}} \log g(z)\,\mathrm{d}H(z,t;\beta)$. Even though H depends upon β, K_1 is a constant, depending only upon the marginal distribution of Z. Now $B(\theta) = E\log f(T;\theta)$ which depends upon the marginal distribution of T. The marginal distribution of T depends on θ via an expression for total probability and the conditional distributions of T given Z. However, writing $f(t) = \int f(t|z)g(z)\,\mathrm{d}z$, then $\partial f(t)/\partial\theta = \int \partial f(t|z)/\partial w(\mathrm{d}w/\mathrm{d}\theta)g(z)\,\mathrm{d}z$, assuming we can interchange the limiting processes. Thus, $\partial f(t)/\partial\theta = E\{Z(\partial f(t|Z)/\partial W)\} = 0$. Under the condition, $B(\theta) = E\log f(T;\theta) = E\log f(T)$ does not depend upon θ and so can be written as a constant K_2. We then have $K = K_1 - K_2$.

The condition indicates, that for each fixed t, the rate of change of the conditional density with respect to $\beta' Z$ is uncorrelated with the covariate Z itself. For normal regression this can be verified. More generally the condition would only be approximately met. □

In view of the above theorem we can view $\hat{\rho}^2_{\mathrm{XOQ}}$ either as; an estimator of the explained randomness in Z given T or, as an approximation to ρ^2, the explained randomness in the ranks of T given Z. This second interpretation is the one that corresponds most closely to that required in the majority of applications. A second theorem leads to further simplification including an approximation that is very easily obtained. First we define

$$\bar{\Gamma}(\hat{\beta}) = 2\sum_{i=1}^{n} W(X_i)\log\left\{\frac{\pi_i(X_i;\hat{\beta})}{\pi_i(X_i;0)}\right\} \tag{12}$$

Again, the right-hand side of the above should be divide by $\sum_1^n W(X_i)$ if it is less than one. We should compare this equation with equation (10). Furthermore we have;

Theorem 3
Assuming the data are generated by equation (1) and that the support for C and T coincides, then $|\bar{\Gamma}(\hat{\beta}) - \hat{\Gamma}(\hat{\beta})|$ converges in probability to zero.

Proof

$$\begin{aligned}
&\log \pi_i(X_i;\hat{\beta}) - \sum_{j=1}^{n} \pi_j(X_i;\hat{\beta}) \log \pi_j(X_i;\hat{\beta}) \\
&= \log \frac{Y_i(X_i)\exp(\hat{\beta}' Z_i(X_i))}{\sum_{l=1}^{n} Y_l(X_i)\exp(\hat{\beta}' Z_l(X_i))} - \sum_{j=1}^{n} \pi_j(X_i;\hat{\beta}) \log \frac{Y_j(X_i)\exp(\hat{\beta}' Z_j(X_i))}{\sum_{l=1}^{n} Y_l(X_i)\exp(\hat{\beta}' Z_l(X_i))} \\
&= Y_i(X_i)\cdot \hat{\beta}' Z_i(X_i) - \sum_{j=1}^{n} \pi_j(X_i;\hat{\beta})\cdot Y_j(X_i)\cdot \hat{\beta}' Z_j(X_i) \\
&= \hat{\beta}' \{Z_i(X_i) - \mathscr{E}(X_i;\hat{\beta})\}
\end{aligned}$$

where $\mathscr{E}(X_i;\hat{\beta}) = \sum_{j=1}^{n} Y_j(X_i) Z_j(X_i) \pi_j(X_i;\hat{\beta})$. So

$$\bar{\Gamma}(\hat{\beta}) - \hat{\Gamma}(\hat{\beta}) = 2\hat{\beta}' \sum_{i=1}^{n} W(X_i)\{Z_i(X_i) - \mathscr{E}(X_i;\hat{\beta})\}$$

which is asymptotically zero [8]. □

Next, if we first define,

$$\bar{\Gamma}_A(\hat{\beta}) = \frac{2}{k} \sum_{i=1}^{n} \delta_i \log \left\{ \frac{\pi_i(X_i;\hat{\beta})}{\pi_i(X_i;0)} \right\} \tag{13}$$

where k is the total number of failures, then we have;

Corollary 1
In the absence of censoring $\bar{\Gamma}_A(\hat{\beta}) = \bar{\Gamma}(\hat{\beta})$.

This is immediate since, in the absence of censoring, $k = n$ and $W(X_i) = W(T_i) = 1/n$. But, more generally, in the presence of independent censoring, we can take $\bar{\Gamma}_A(\hat{\beta})$ to be a good approximation to $\bar{\Gamma}(\hat{\beta})$. Both $\bar{\Gamma}_A(\hat{\beta})$ and $\bar{\Gamma}(\hat{\beta})$ represent empirical expectations of the same quantity, the difference being in how we assign the total probability mass of one. For most observed censoring patterns, we would not anticipate seeing much discrepancy between $\bar{\Gamma}_A(\hat{\beta})$ and $\bar{\Gamma}(\hat{\beta})$. Since $\bar{\Gamma}_A(\hat{\beta})$ is particularly straightforward to evaluate, using a simple transformation of the partial likelihood ratio statistic, we would recommend it for routine use. In the absence of censoring, note that $\hat{\rho}^2(\hat{\beta})$, based upon $\bar{\Gamma}_A(\hat{\beta})$, coincides with the correlation measure provided by a SAS survival analysis book [10]. The Allison measure, however, uses n in place of k in (12), therefore will depend upon an independent censoring mechanism regardless of population effects. In particular it approaches the value zero as the percentage of censored observations approaches one. On the other hand, for all independent censoring mechanisms, $\hat{\rho}^2(\hat{\beta})$ approaches a population equivalent, this equivalent being close to $\rho^2(\beta)$, and therefore interpretable as a percentage of explained randomness. Furthermore, if 'being close' is not good enough and we insist upon a consistent estimate of $\rho^2(\beta)$ then this requires little in the way of extra work.

4. SIMULATION

In this section we compare the aforementioned measures of explained randomness via simulation, for different strength of regression effects, different censoring percentages and different covariate distributions. Data are simulated with hazard function $\lambda(t) = \exp(-\beta Z)$. The distributions of Z are standardized to have the same variances. The censoring mechanism is uniform $[0, \tau]$. Table I contains the results of a large sample ($n = 5000$) comparison. The last three columns of the table is from Xu and O'Quigley [3], where ρ^2_W and $\rho^2_{W,A}$ are defined in Reference [4]. The coefficient ρ^2_W is their sample-based estimate, and $\rho^2_{W,A}$ a simpler approximation to ρ^2_W. The measure denoted ρ^2_n is from Allison [10], where the partial likelihood ratio statistic is incorrectly divided by the sample size n instead of the number of events k.

From the table we see that the measure ρ^2_n decreases dramatically with the percentage of censoring, which is consistent with our earlier discussion. The measures ρ^2 and ρ^2_{XOQ} have close agreement, as they are asymptotically equivalent. The measure ρ^2_k appears to be a good approximation to ρ^2, even in the presence of heavy censoring. The measures from Kent and O'Quigley [4] also have good agreement with these three measures.

Table I. A simulated comparison of the measures ($n = 5000$).

$\exp(\beta)$	Per cent censored	Covariate	ρ^2_n	ρ^2_k	ρ^2	ρ^2_{XOQ}	ρ^2_W	$\rho^2_{W,A}$
2	0	c	0.103	0.103	0.103	0.102	0.096	0.119
	50	c	0.053	0.106	0.097	0.108	0.089	0.122
	90	c	0.013	0.117	0.107	0.105	0.103	0.099
	0	d	0.094	0.094	0.094	0.102	0.113	0.118
	50	d	0.060	0.113	0.086	0.110	0.114	0.121
	90	d	0.012	0.114	0.116	0.106	0.125	0.100
4	0	c	0.293	0.293	0.293	0.295	0.304	0.338
	50	c	0.197	0.351	0.356	0.334	0.298	0.344
	90	c	0.048	0.390	0.366	0.340	0.279	0.342
16	0	c	0.603	0.603	0.603	0.598	0.623	0.664
	50	c	0.473	0.713	0.679	0.690	0.622	0.668
	90	c	0.116	0.725	0.765	0.723	0.605	0.670
64	0	c	0.757	0.757	0.757	0.758	0.785	0.815
	50	c	0.641	0.873	0.857	0.848	0.790	0.815
	90	c	0.197	0.865	0.886	0.876	0.763	0.816
	0	d	0.679	0.679	0.679	0.681	0.777	0.814
	50	d	0.633	0.865	0.859	0.860	0.776	0.815
	90	d	0.128	0.726	0.742	0.756	0.795	0.792

ρ^2_W, $\rho^2_{W,A}$: [4].
c: continuous Z-Uniform $(0, \sqrt{3})$.
d: dichotomous Z-0,1 with equal probabilities.

Table II. A simulated comparison of the measures ($n = 100$).

$\exp(\beta)$	Per cent censored	Covariate	ρ_n^2	ρ_k^2	ρ^2	ρ_{nc}^2
2	0	c	0.105 (0.053)	0.105 (0.053)	0.105 (0.053)	0.105 (0.053)
	50	c	0.059 (0.041)	0.114 (0.079)	0.106 (0.084)	0.097 (0.050)
	90	c	0.021 (0.024)	0.173 (0.186)	0.149 (0.279)	0.106 (0.051)
	0	d	0.103 (0.056)	0.103 (0.056)	0.103 (0.056)	0.103 (0.056)
	50	d	0.065 (0.045)	0.122 (0.082)	0.118 (0.090)	0.106 (0.054)
	90	d	0.022 (0.027)	0.180 (0.195)	0.178 (0.224)	0.107 (0.051)
4	0	c	0.288 (0.071)	0.288 (0.071)	0.288 (0.071)	0.288 (0.071)
	50	c	0.186 (0.062)	0.337 (0.105)	0.319 (0.115)	0.291 (0.066)
	90	c	0.047 (0.038)	0.357 (0.231)	0.349 (0.271)	0.292 (0.066)
16	0	c	0.587 (0.056)	0.587 (0.056)	0.587 (0.056)	0.587 (0.056)
	50	c	0.443 (0.066)	0.686 (0.079)	0.662 (0.093)	0.586 (0.055)
	90	c	0.118 (0.054)	0.695 (0.177)	0.658 (0.274)	0.591 (0.055)
64	0	c	0.750 (0.033)	0.750 (0.033)	0.750 (0.033)	0.750 (0.033)
	50	c	0.616 (0.048)	0.847 (0.043)	0.827 (0.057)	0.743 (0.037)
	90	c	0.182 (0.062)	0.857 (0.097)	0.851 (0.120)	0.751 (0.037)
	0	d	0.676 (0.028)	0.676 (0.028)	0.676 (0.028)	0.676 (0.028)
	50	d	0.627 (0.036)	0.857 (0.037)	0.846 (0.042)	0.673 (0.028)
	90	d	0.135 (0.042)	0.750 (0.090)	0.747 (0.120)	0.676 (0.030)

Next we carried out simulations to evaluate the finite sample ($n = 100$) behaviours of ρ_n^2, ρ_k^2 and ρ^2 (Table II). Since the difference among these three measures lies in their ways of handling censorship, for comparison purposes in the last column of the table we also provide the value of the measure for the same simulated data without censoring, ρ_{nc}^2. In the parenthesis are the standard errors from the 200 simulations. As we can see ρ_n^2 again behaves quite poorly with increasing censoring. When compared to ρ_{nc}^2, ρ^2 is slightly less biased than ρ_k^2, but ρ_k^2 turns out to have smaller standard errors than ρ^2. We also computed the mean squared errors (MSE), and although not shown here, the MSE for ρ_k^2 is generally slightly smaller than ρ^2.

5. ILLUSTRATION

The first example concerns the well-known Freireich data, used and described in the founding paper of Cox [1] as well as in all of the available texts on survival analysis. The study recorded the remission times of 42 patients with acute leukaemia treated by 6-mercaptopurine (6-MP) or placebo. A visual inspection of the separation afforded by the two Kaplan–Meier curves, corresponding to the two prognostic groups, is, of itself, enough to suggest that there are quite important predictive effects. The regression coefficient was estimated as $\hat{\beta} = 1.65$. We find that $\hat{\rho}^2 = 0.37$ and that $\hat{\rho}_{\mathrm{XOQ}}^2 = 0.40$, indicating that around some 40 per cent of the

randomness can be explained by the prognostic factor treatment. The much simpler calculation, based on the modification of the coefficient introduced by Allison [10], results in the value 0.42 giving, as we would expect, quite close agreement. The uncorrected coefficient of Allison results in the slightly depressed value 0.32 which is not all that far removed from the others but this is simply due to the fact that, for these data, only 12 out of the 42 observations were censored. As the censoring diminishes the corrected and uncorrected Allison measures converge and, ultimately, when there is no censoring, they are identical. A second example comes from a study of 2174 breast cancer patients, followed over a period of 15 years at the Institut Curie in Paris, France. A large number of potential and known prognostic factors were recorded. The most important prognostic factor among those anticipated as having some prognostic importance was stage. Measuring, not so indirectly, the evolution of the disease our intuition tells us that we would expect to record a high degree of explained randomness from this covariate. Applying the uncorrected Allison coefficient in the standard SAS program we find an estimated explained randomness of 7 per cent. This appears rather low given what we know and, indeed, if we calculate the corrected coefficient, as described above, we then find an estimated 27 per cent of the randomness in survival explained by stage. This is relatively high but does correspond much more closely to a figure we might expect. In addition, even after adjusting for other important prognostic factors histology grade, progesterone receptor status and tumor size, stage still explain an estimated 7 per cent of randomness in survival. This estimate is easily obtained from calculating the corrected coefficients for the full model with all the four covariates which gives 0.42, and for the reduced model with just grade, receptor status and tumor size, which gives 0.38, then applying equation (6).

6. CONCLUSION

The motivation behind the development of the Xu and O'Quigley [3] coefficient of explained randomness was to simplify inference and enable the exploitation of routine calculations in proportional hazards regression. At the same time further generality was achieved such as the ability to include time-dependent covariates in the analysis. It appeared that the price to pay for these gains was essentially in interpretation, the coefficient ρ^2_{XOQ} being less readily interpretable than ρ^2. However, in the light of Theorem 2, this difficulty largely disappears. The useful interpretation of ρ^2 holds equally well for ρ^2_{XOQ}, providing that we only wish to explain randomness in the ranking of realizations of T. In view of the more straightforward, and well established, inferential procedures available for ρ^2_{XOQ}, in addition to its capacity to accommodate time-dependent covariates, we propose that the coefficient of Xu and O'Quigley replace that originally developed by Kent and O'Quigley. And, as a final very simple take-home message, we point out that a computationally trivial modification of the coefficient suggested by the SAS book, replacing a denominator of size n by one of size k, produces a biased estimate of ρ^2_{XOQ}. We would expect the bias to be small, in most applications an order or orders of magnitude smaller than the standard error of estimate. The modified SAS coefficient can then be given an interpretation as one of explained randomness and could be adopted for routine use.

REFERENCES

1. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* 1972; **34**:187–220.
2. Cox DR. Partial likelihood. *Biometrika* 1975; **63**:269–276.
3. Xu R, O'Quigley J. A measure of dependence for proportional hazards models. *Journal of Nonparametric Statistics* 1999; **12**:83–107.
4. Kent JT, O'Quigley J. Measures of dependence for censored survival data. *Biometrika* 1988; **75**:525–534.
5. Kent JT. Information gain and a general measure of correlation. *Biometrika* 1983; **70**(1):163–174.
6. Kent JT. The underlying structure of nonnested hypothesis tests. *Biometrika* 1986; **73**(2):333–344.
7. Heinzl H, Stare J. Using SAS to calculate the Kent and O'Quigley measure of dependence for the Cox proportional hazards model. *Computer Methods and Programs in Biomedicine* 2000; **63**:71–76.
8. Xu R, O'Quigley J. Estimating average regression effect under non-proportional hazards. *Biostatistics* 2000; **1**:423–439.
9. Xu R, O'Quigley J. Proportional hazards estimate of the conditional survival function. *Journal of the Royal Statistical Society, Series B* 2000; **62**:667–680.
10. Allison PD. *Survival Analysis Using the SAS System*. SAS Institute Inc.: Cary, North Carolina, U.S.A., 1995.
11. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 1982; **10**:1110–1120.
12. O'Quigley J, Flandre P. Predictive capability of proportional hazards regression. *Proceedings of the National Academy of Sciences of the United States of America* 1994; **91**:2310–2314.
13. Xu R. Inference for the proportional hazards model. *Ph.D. Thesis*, University of California, San Diego, 1996.

