



HAL
open science

Accuracy of predictive ability measures for survival models

Philippe Flandre, Reena Deutsch, John O'Quigley

► **To cite this version:**

Philippe Flandre, Reena Deutsch, John O'Quigley. Accuracy of predictive ability measures for survival models. *Statistics in Medicine*, 2017, 10.1002/sim.7342 . hal-01552992

HAL Id: hal-01552992

<https://hal.sorbonne-universite.fr/hal-01552992v1>

Submitted on 3 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accuracy of predictive ability measures for survival models

Philippe Flandre,^{a,*†}  Reena Deutsch^b and John O'Quigley^c

One aspect of an analysis of survival data based on the proportional hazards model that has been receiving increasing attention is that of the predictive ability or explained variation of the model. A number of contending measures have been suggested, including one measure, $R^2(\beta)$, which has been proposed given its several desirable properties, including its capacity to accommodate time-dependent covariates, a major feature of the model and one that gives rise to great generality. A thorough study of the properties of available measures, including the aforementioned measure, has been carried out recently. In that work, the authors used bootstrap techniques, particularly complex in the setting of censored data, in order to obtain estimates of precision. The motivation of this work is to provide analytical expressions of precision, in particular confidence interval estimates for $R^2(\beta)$. We use Taylor series approximations with and without local linearizing transforms. We also consider a very simple expression based on the Fisher's transformation. This latter approach has two great advantages. It is very easy and quick to calculate, and secondly, it can be obtained for any of the methods given in the recent review. A large simulation study is carried out to investigate the properties of the different methods. Finally, three well-known datasets in breast cancer, lymphoma and lung cancer research are given as illustrations. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: predictive ability; prognostic factors; proportional hazards model; residuals; survival analysis

1. Introduction

Whereas the use of the Cox regression model in the analysis of censored survival data has become routine, the corresponding use of R^2 -type measures of explained variation is still not that common. Unlike the case of the linear model where there is essentially only a single contender, for survival models, there is more than one possibility. In an extensive review, Choodari-Oskooei et al [1, 2] examine the properties of a number of possible measures of R^2 for survival data. Our purpose here is to add a piece to that development; this piece being the provision of a variance estimate for all of the measures discussed by [1]. One of our suggestions is rough and ready in that it is very easily calculated – coming under the heading of those techniques deemed desirable by student, techniques that ‘can be calculated on the back of a railway ticket while waiting for the train’ – but, as the simulations show, it may be too rough when the purpose is to obtain a reasonably accurate test. Nonetheless, agreement with the bootstrap calculations of [1] is generally good, and our suggestion is that it be used to obtain a rough interval estimate of the values compatible with an estimated R^2 . The general motivation for this variance estimate comes from the known approximately pivotal property of the Fisher's transformation. Apart from this pivotal property, there is no other obvious reason why this coefficient should work so well, unless we take a linear model as a first approximation to the nonlinear Cox model, and it is a somewhat surprising result.

Aside from the Fisher transformation, we investigate analytical approximations for the variance of the R^2 coefficients for survival data based on the proposals of [3]. There are two possible approaches here; the first makes use of the monotonicity relationship between the conditional expectation of an explained variation measure and the regression coefficient, β , and the second appeals to large sample approximations.

^aSorbonne Universités, UPMC Univ Paris 06, INSERM, Institut Pierre Louis d'épidémiologie et de Santé Publique (IPLESP UMR-S 1136), F75013 Paris, France

^bDepartment of Psychiatry, University of California at San Diego, La Jolla, CA 92093, U.S.A.

^cLaboratoire de Statistique Théorique et Appliquée, UPMC, Paris, France

*Correspondence to: Philippe Flandre, Sorbonne Universités, UPMC Univ Paris 06, INSERM, Institut Pierre Louis d'épidémiologie et de Santé Publique (IPLESP UMR-S 1136) F75013 Paris, France.

†E-mail: philippe.flandre@iplesp.upmc.fr

Different transformations, aimed at local linearization, are considered. Local linearization corresponds to taking the first two terms of a Taylor expansion of R^2 around the maximum likelihood estimate, $R^2(\hat{\beta})$. Because such an expansion is available to us for any monotonic transformation of R^2 , we investigate some specific cases because some may linearize more successfully (more rapidly with sample size) than others. It is interesting to note good agreement between the different approaches and, also, good agreement with the results of [1] based on their use of the bootstrap. Somewhat like R^2 itself, the goal is to give more an impression of overall strength of effect rather than something on which we might wish to base accurate error controlled decision-making.

Values of R^2 close to zero may still correspond to effects of interest, but we know that from the standpoint of individual prediction, there is much more noise than signal. Conversely, larger values may be such that the endeavour of building prognostic models that can be used effectively in practice becomes a plausible one. Thus, it would be helpful to provide interval estimates of R^2 or to at least be able to provide a reasonable range within which we might feel the actual value of predictive effect can be found. Note that such an interpretation would be close in spirit to a Bayesian one, rather than a more classical confidence interval interpretation, but, again, the goal is not to be so refined. It is more to have some rough and ready tools of service to us in our quest to build useful predictive models.

In the following section, we recall the basic structure of the main measure of our focus in this paper. The approximations, based on Taylor series, rely upon this structure and can therefore be seen as being specific to this measure. The Fisher transform on the other hand is very general and would apply to any measure, although we only study its performance in relation to the measure given in this section. Section 3 has two parts: the first part is devoted to Taylor series approximations of R^2 , as defined in Section 2, and some simple monotonic transformations of R^2 . The second part of Section 3 provides motivation and a simple expression for the variance of an invertible function of R^2 . The expression depends only upon R^2 , for which we use a plug-in estimate and the number of observed failure times. Section 4 presents a simulation to investigate coverage properties and how they might be influence by the size of the true population equivalent of R^2 , the rate of censoring, the distribution of the covariates and the correlation of the covariates in some relatively simple cases. Some real examples are considered in Section 5, and some further points of discussion are raised in Section 6.

2. R^2 measures for proportional hazards models

Let T denote the failure time, C the censoring time and $Z(t)$ is $p \times 1$ vector of explanatory variables. Assume that T and C are independent conditional on $Z(t)$. Suppose the data consist of n independent replicates of $(X, \Delta, Z(t))$, where $X = \min(T, C)$, $\Delta = I(T \leq C)$, 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 . We will also use the counting process notation: let $N_i(t) = I\{T_i \leq t, T_i \leq C_i\}$ and $\tilde{N}(t) = \sum_{i=1}^n N_i(t)$. The proportional hazards model [4] specifies that the hazard function for the failure time T associated with $Z(t)$ takes the form

$$\lambda(t; Z(t)) = \lambda_0(t) \exp(\beta' Z(t)), \quad (2.1)$$

where $\lambda_0(\cdot)$ is an unspecified hazard function and β is a $p \times 1$ vector of unknown regression parameters. The usual estimate, $\hat{\beta}$, of β is found by solving $U(\beta) = 0$, where $U(\beta)$ is the score statistic of derivatives of the logarithm partial likelihood $U(\beta) = \sum_{i=1}^n \Delta_i r_i(\beta)$ where $r_i(\beta)$ are the partial Schoenfeld's residuals at each observed failure of the time T_i , $r_i(\beta) = Z_i(t) - \mathcal{E}_\beta[Z|t]$ with

$$\mathcal{E}_\beta[Z|t] = \sum_{j=1}^n Z_j(t) \pi_j(\beta, t), \quad \pi_j(\beta, t) = \frac{Y_j(t) \exp\{\beta' Z_j(t)\}}{\sum_{i=1}^n Y_i(t) \exp\{\beta' Z_i(t)\}}, \quad (2.2)$$

the quantity $\pi_j(\beta, t)$ being interpretable as the conditional probability that at time t , it is precisely individual i who is chosen to fail, given all the individuals at risk and given that one failure occurs [3, 5]. When $\beta = 0$, $\pi_i(0, t)$ is simply the empirical distribution of the covariate given t , assigning equal weight to each sample subject in the risk set. In other words, given t and given the subjects in the risk set at t , we have a sampling model defined on this set by which every individual has the same probability of being selected.

Schoenfeld residuals of the proportional hazards model [6] measure the discrepancy between the observed value of the covariate $Z_i(t)$, from the individual i who fails at time t , and its expected value under

the model and that a single selection is being made from those subjects at risk. In the multivariate model (2.1), the dependence of the survival time variable on the covariates is via the prognostic index [5, 7]

$$\eta(t) = \beta'Z(t).$$

Two individuals with different Z 's but the same η should have the same survival probability. Let us define the quantity $\mathcal{I}(b)$ for $b = 0, \beta$ by

$$\mathcal{I}(b) = \sum_{i=1}^n \int_0^{\infty} \{\eta_i(t) - \beta' \mathcal{E}_b(Z|t)\}^2 dN_i(t). \quad (2.3)$$

The measure of predictive ability of a proportional hazards model was therefore defined by [5]

$$R^2(\beta) = 1 - \mathcal{I}^{-1}(0)\mathcal{I}(\beta), \quad (2.4)$$

where in practice, β is replaced by $\hat{\beta}$. $R^2(\hat{\beta})$ is a sample-based quantity which, under the usual asymptotic framework [7], converges to a population counterpart denoted $\Omega^2(\beta)$ (details are outlined in O'Quigley and Xu where the properties of the coefficient were investigated). The ratio $\mathcal{I}(\hat{\beta})/\mathcal{I}(0)$ will approach 1 when there is almost no predictive effect and decrease towards the value 0 with increasing effect. We can generalize Schoenfeld's residuals in the multivariate settings where for each observed failure time T_i , residuals are defined by $w_i(\beta) = \{\eta_i(t) - \beta' \mathcal{E}_b(Z|t)\}$. Then, in the case of a single covariate in the model, residuals are a function of the Schoenfeld residuals with $w_i(\beta) = \beta r_i(\beta)$ and the predictive index in (2.4) reduced to the measure introduced by [3]

$$R^2(\hat{\beta}) = 1 - \frac{\sum_{i=1}^n \Delta_i r_i^2(\hat{\beta})}{\sum_{i=1}^n \Delta_i r_i^2(0)}, \quad (2.5)$$

where $\Delta_i = I(T_i \leq C_i)$. Therefore, predictability or explained variability, in the context of proportional hazards regression, refers to predictability or variability of the value of the covariate, conditional upon the failure time and subjects at risk at that time. This approach has the great advantage of allowing us to consider the potential inclusion of time-dependent variables, which is one of the model's most attractive features. The use of time-dependent variables not only is a key property of the model but also corresponds to many applied problems in chronic illnesses such as cancer and HIV. Although, in the context of survival data, it does seem more natural to want predictability to refer to time, because of the features of the model, in particular the importance of the residuals based on the value of the covariate, our approach is intuitively logical and has desirable properties [3, 5]. The exponentially tilted distribution given by equation (2.2) puts increasing weight on the higher values as the coefficient β increases. When β assumes increasingly large negative values, then the same occurs on the smaller values. For a single binary explanatory variable, as β increases without bound, then all of the weight is put on the covariate value one. If we have perfect discrimination and the ones are systematically selected until they are exhausted, then as $\beta \rightarrow \infty$, $R^2(\beta) \rightarrow 1$. If the zeros are selected until they are exhausted then, again, as $\beta \rightarrow \infty$, $R^2(\beta) \rightarrow 1$. For a covariate taking on several values, or continuous, then, if at each observed failure time, the value of the covariate of the individual i who fail at T_i is the smallest (greatest) value among the sample values in the risk set, then the prediction is once more perfect and $R^2(\beta) = 1$. Given the covariate, $R^2(\beta)$ increases monotonically in $|\beta|$ so that increasing strength of effect translates and increasing values of $R^2(\beta)$.

3. Confidence interval estimates of $R^2(\beta)$

3.1. Approximations based on linearization

For convenience, we denote $R^2(\beta)$ by R^2 . R^2 is a function of random variables $Z_1(t), \dots, Z_k(t)$, where k is the number of failures. We use Taylor series methods by expanding R^2 through a series expansion to the first order so that we can obtain an estimate of $\text{Var}(R^2)$. For $R^2 = f(\beta'Z_1(t), \dots, \beta'Z_k(t))$ and letting $w_{i\beta} = w_i(\beta)$, $w_{i0} = w_i(0)$, and $\eta_i = \eta_i(t)$, $i = 1, \dots, k$,

$$\text{Var}(R^2) \approx \sum_{i=1}^k \left\{ \frac{\partial R^2}{\partial \eta_i} \right\}^2 \text{Var}(\eta_i) = \sum_{i=1}^k \left\{ \frac{\partial R^2}{\partial Z_{ij}} / \frac{\partial \eta_i}{\partial Z_{ij}} \right\}^2 \text{Var}(\eta_i), \quad (3.6)$$

where $j = 1, \dots, p$, where p is the dimension of the parameter space. These calculations are straightforward and, after gathering terms, we have that

$$\begin{aligned} \text{Var}(R^2) &= \sum_{i=1}^k \left[\frac{2\Delta_i w_{i\beta} \sum \Delta_j w_{j0}^2 - 2\Delta_i w_{i0} \sum \Delta_j w_{j\beta}^2}{\left(\sum \Delta_j w_{j0}^2\right)^2} \right]^2 \text{Var}(\eta_i) \\ &= \frac{4}{\left(\sum \Delta_j w_{j0}^2\right)^4} \sum_{i=1}^k \Delta_i^2 \left[w_{i\beta} \sum \Delta_j w_{j0}^2 - w_{i0} \sum \Delta_j w_{j\beta}^2 \right]^2 \text{Var}(\eta_i), \end{aligned} \quad (3.7)$$

where $\text{Var}(\eta_i)$ is approximated by $\sum_{j=1}^p \beta_j^2 \text{Var}(Z_{ij}(t)) + 2 \sum_{j < \ell} \beta_j \beta_\ell \text{Cov}(Z_{ij}(t), Z_{i\ell}(t))$, $\text{Var}(Z_{ij}(t)) = \mathcal{E}_\beta[Z_j^2|t] - \mathcal{E}_\beta^2[Z_j|t]$, $j = 1, \dots, p$ and $\text{Cov}(Z_{ij}(t), Z_{i\ell}(t)) = \mathcal{E}_\beta[Z_j Z_\ell|t] - \mathcal{E}_\beta[Z_j|t] \mathcal{E}_\beta[Z_\ell|t]$, $j < \ell$. Furthermore letting $S = \log(1 - R^2)$,

$$\text{Var} S \approx \sum_{i=1}^k \left\{ \frac{\partial \log(1 - R^2)}{\partial \eta_i} \right\}^2 \text{Var}(\eta_i) = \sum_{i=1}^k \left\{ \frac{\partial \log(1 - R^2)}{\partial Z_{ij}} / \frac{\partial \eta_i}{\partial Z_{ij}} \right\}^2 \text{Var}(\eta_i),$$

where $j = 1, \dots, p$. Again, gathering terms, we have that

$$\begin{aligned} \text{Var}[\log(1 - R^2)] &= \sum_{i=1}^k \left[2\Delta_i \left(\frac{w_{i\beta}}{\sum \Delta_j w_{j\beta}^2} - \frac{w_{i0}}{\sum \Delta_j w_{j0}^2} \right) \right]^2 \text{Var}(\eta_i) \\ &= \frac{4}{\left(\sum \Delta_j w_{j0}^2 \sum \Delta_j w_{j\beta}^2\right)^2} \sum_{i=1}^k \Delta_i^2 \left[w_{i\beta} \sum \Delta_j w_{j0}^2 - w_{i0} \sum \Delta_j w_{j\beta}^2 \right]^2 \text{Var}(\eta_i), \end{aligned}$$

In the same way, letting $S^* = \log(R^2/(1 - R^2))$, and regrouping terms, the variance of S^* is approximated as

$$\begin{aligned} \text{Var}(S^*) &\approx \sum_{i=1}^k \left[2\Delta_i \left(\frac{(w_{i0} - w_{i\beta})}{\sum \Delta_j w_{j0}^2 - \sum \Delta_j w_{j\beta}^2} - \frac{w_{i\beta}}{\sum \Delta_j w_{j\beta}^2} \right) \right]^2 \text{Var}(\eta_i) \\ &= \frac{4}{\left(\left(\sum \Delta_j w_{j0}^2 - \sum \Delta_j w_{j\beta}^2\right) \sum \Delta_j w_{j\beta}^2\right)^2} \\ &\quad \times \sum_{i=1}^k \Delta_i^2 \left[(w_{i0} - w_{i\beta}) \sum \Delta_j w_{j\beta}^2 - w_{i\beta} \left(\sum \Delta_j w_{j0}^2 - \sum \Delta_j w_{j\beta}^2\right) \right]^2 \text{Var}(\eta_i). \end{aligned}$$

We first obtain a 95% confidence interval $[S_L, S_U]$ and $[S_L^*, S_U^*]$ for S and S^* , respectively. Then, the endpoints of these intervals were inserted into the transformed confidence interval for R^2 , $[1 - e^{S_U}, 1 - e^{S_L}]$ and $[e^{S_L^*}/(1 + e^{S_L^*}), e^{S_U^*}/(1 + e^{S_U^*})]$, by virtue of the monotonicity of S and S^* , respectively, as a function of R^2 .

3.2. Fisher's transformation

Suppose we were to make a bivariate Edgeworth approximation to the joint distribution of T and Z . Obviously, we would need to make many more assumptions, in particular some parametric joint model for T and Z , but in fact, we will only need suppose that such a distribution exists. We will not carry out the calculation of the terms needed in the Edgeworth expansion, but we will suppose that they could be obtained if we had the joint model. The purpose of this is purely motivational, and in practice, the approximation we use is particularly simple and does not need any further calculation [8].

We take

$$\hat{\theta} = \tanh^{-1} R = \frac{1}{2} \log \left(\frac{1 + R}{1 - R} \right),$$

and, in the population, $\theta = \tanh^{-1} \Omega$. Following the development of [9], it is possible to develop R as an infinite polynomial series in $\hat{\theta} - \theta$. We find, for $a = \hat{\theta} - \theta$,

$$\begin{aligned} f(a) = & \frac{nE(\Delta) - 2}{\sqrt{2\pi(nE(\Delta) - 1)}} \left[1 + \frac{a\theta}{2} + \left(\frac{2 + \theta}{8(nE(\Delta) - 1)} + \frac{4 - \theta^2}{8} a^2 + \frac{nE(\Delta) - 1}{12} a^4 \right) \right. \\ & + \left(\frac{4 - \theta^2}{16(nE(\Delta) - 1)} + \frac{4 + 3\theta^2}{48} a^2 + \frac{n - 1}{24} a^4 \right) a\theta \\ & + \left(\frac{4 + 12\theta^2 + 9\theta^4}{128(nE(\Delta) - 1)^2} + \frac{8 - 2\theta^2 + 3\theta^4}{64(nE(\Delta) - 1)} a^2 + \frac{8 + 4\theta^2 - 5\theta^4}{128} a^4 \right. \\ & \left. \left. + (nE(\Delta) - 1) \frac{28 - 15\theta^2}{1440} a^6 \right) \right] \exp -(nE(\Delta) - 1)a^2/2. \end{aligned}$$

Furthermore, neglecting powers of n higher than the third, for the first two moments, and higher than the fourth for the next two, we have

$$E(a) = \frac{\Omega}{2(nE(\Delta) - 1)} \left[1 + \frac{5 + \Omega^2}{4(nE(\Delta) - 1)} + \frac{11 + 2\Omega^2 + 3\Omega^4}{8(nE(\Delta) - 1)^2} \right] \quad (3.8)$$

$$\text{Var}(a) = \frac{1}{nE(\Delta) - 1} \left[1 + \frac{4 - \Omega^2}{2(nE(\Delta) - 1)} + \frac{22 - 6\Omega^2 - 3\Omega^4}{6(nE(\Delta) - 1)^2} \right]. \quad (3.9)$$

Given this result, we can calculate the first four moments, leading to

$$\text{Var}(\theta - \tanh^{-1} \Omega) = \frac{1}{nE(\Delta) - 1} + \frac{4 - \theta^2}{2(nE(\Delta) - 1)^2}, \quad (3.10)$$

which, in turn, can be approximated by $(nE(\Delta) - 1)^{-1} + 2(nE(\Delta) - 1)^{-2} = 1/(nE(\Delta) - 3)$. In practice, we do not know the value of $E(\Delta)$. We can though, under an independent censoring assumption, replace it by $\hat{\Delta}$, defined as the proportion of failures in the sample scaled to the interval (0,1). This is once more a working assumption. In total then, for this rough and ready approximation, there have been a string of working assumptions. It is not easy to theoretically gauge the impact of these approximations on the usefulness of the final result and, so, rather than do that we simply investigated the performance of this approximation via a range of simulations. Also, of course, we were able to compare the results of the rough and ready approximation to those obtained via Taylor series expansion and the level of agreement was very satisfactory. Once we have the above intervals, we transform them back to the original scale for R^2 .

Asymptotically, $E(\theta) = \tanh^{-1} \rho$ and, approximately, $\text{var}(\theta) = 1/(k - 3)$, where k is the number of deaths. We first obtain, a 95% confidence interval $[\theta_L, \theta_U]$ and then a 95% confidence interval for r , with $r = \sqrt{R^2(\hat{\beta})}$, $[(e^{2\theta_L} - 1)/e^{2\theta_L} + 1, (e^{2\theta_U} - 1)/e^{2\theta_U} + 1]$. These intervals will not be symmetric, but they are guaranteed to lie within the interval (0,1).

4. Simulations

A simulation study was conducted to evaluate the performance of the confidence intervals obtained using both Taylor series and Fisher's transformation. For each simulation, a sample size of $n = 100$ or $n = 400$ was generated from an exponential distribution with parameter $\lambda = \exp(\beta_0 + \beta z)$. The covariate z was a binary covariate either normally distributed or uniformly distributed with in all cases $E[z] = 0.5$ and $\text{Var}[z] = 0.25$. For example, for a binary covariate, $z = 0$ with probability 0.5 and $z = 1$ with probability 0.5 was generated. We investigated absence of censoring, and presence of 25%, 50% and 75% of censoring. The 95% confidence intervals of $R^2(\hat{\beta})$ were calculated using all the methods presented previously, and the true value $\Omega^2(\beta)$ was estimated as described in [3]. Five thousand replications of such samples were generated, and the coverage probability of each 95% confidence interval was calculated. A further simulation considered the case of bivariate normal covariables. The means and variances were the same as for the case of a single variable, and the correlation parameter took the values: 0.0, 0.3, 0.6 and

Table I. Simulation confidence intervals for R^2 with a binary covariate based on Taylor series methods and Fisher's transformation.

| Method | $N = 100$ | | | | $N = 400$ | | | |
|--|-------------------------|-------|-------|-------|-------------------------|-------|-------|-------|
| | Percentage of censoring | | | | Percentage of censoring | | | |
| | 0% | 25% | 50% | 75% | 0% | 25% | 50% | 75% |
| $\beta = 1 \quad \mathcal{E}[R^2(\beta)] = 0.19$ | | | | | | | | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 94.7% | 93.7% | 92.0% | 86.5% | 97.3% | 95.9% | 95.4% | 93.3% |
| Based on $\log[1 - R^2]$ | 94.5% | 93.8% | 92.2% | 87.5% | 97.2% | 95.8% | 95.2% | 93.5% |
| Based on $\log[R^2/(1 - R^2)]$ | 97.8% | 97.0% | 96.4% | 94.8% | 97.5% | 97.0% | 96.7% | 97.1% |
| Fisher's transformation | 95.8% | 95.5% | 94.9% | 94.7% | 96.9% | 95.9% | 95.3% | 94.9% |
| $\beta = 2 \quad \mathcal{E}[R^2(\beta)] = 0.49$ | | | | | | | | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 94.4% | 93.6% | 91.5% | 88.1% | 95.3% | 94.8% | 94.4% | 92.8% |
| Based on $\log[1 - R^2]$ | 95.2% | 94.7% | 93.4% | 91.1% | 95.5% | 95.1% | 94.6% | 93.5% |
| Based on $\log[R^2/(1 - R^2)]$ | 95.8% | 95.9% | 95.6% | 95.3% | 95.6% | 95.4% | 95.3% | 94.7% |
| Fisher's transformation | 90.3% | 89.3% | 88.6% | 86.6% | 90.9% | 90.8% | 89.7% | 89.1% |
| $\beta = 3 \quad \mathcal{E}[R^2(\beta)] = 0.73$ | | | | | | | | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 88.4% | 87.8% | 85.8% | 83.6% | 91.3% | 90.3% | 90.2% | 88.3% |
| Based on $\log[1 - R^2]$ | 89.9% | 90.3% | 90.0% | 92.6% | 91.5% | 90.7% | 91.0% | 90.8% |
| Based on $\log[R^2/(1 - R^2)]$ | 90.5% | 91.4% | 91.0% | 87.3% | 91.7% | 91.1% | 91.7% | 91.4% |
| Fisher's transformation | 75.7% | 74.8% | 74.2% | 74.0% | 79.1% | 77.0% | 76.7% | 76.3% |

Table II. Simulation confidence intervals for R^2 with a covariate normally distributed based on Taylor series methods and Fisher's transformation.

| Method | $N = 100$ | | | | $N = 400$ | | | |
|--|-------------------------|-------|-------|-------|-------------------------|-------|-------|-------|
| | Percentage of censoring | | | | Percentage of censoring | | | |
| | 0% | 25% | 50% | 75% | 0% | 25% | 50% | 75% |
| $\beta = 1 \quad \mathcal{E}[R^2(\beta)] = 0.19$ | | | | | | | | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 94.2% | 92.4% | 90.2% | 85.4% | 95.4% | 95.8% | 94.7% | 92.8% |
| Based on $\log[1 - R^2]$ | 94.2% | 92.2% | 90.5% | 86.0% | 95.5% | 95.9% | 94.4% | 93.4% |
| Based on $\log[R^2/(1 - R^2)]$ | 97.0% | 96.7% | 96.2% | 94.4% | 96.3% | 96.9% | 96.6% | 96.4% |
| Fisher's transformation | 95.5% | 95.0% | 94.5% | 94.6% | 95.5% | 96.2% | 95.3% | 94.8% |
| $\beta = 2 \quad \mathcal{E}[R^2(\beta)] = 0.44$ | | | | | | | | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 94.0% | 93.5% | 93.1% | 89.6% | 95.8% | 95.3% | 94.7% | 93.7% |
| Based on $\log[1 - R^2]$ | 94.8% | 94.5% | 93.8% | 91.6% | 95.8% | 95.5% | 94.9% | 94.3% |
| Based on $\log[R^2/(1 - R^2)]$ | 95.5% | 95.6% | 96.0% | 95.3% | 96.2% | 95.7% | 95.6% | 95.4% |
| Fisher's transformation | 93.1% | 92.7% | 93.2% | 91.9% | 94.1% | 93.9% | 93.5% | 93.1% |
| $\beta = 3 \quad \mathcal{E}[R^2(\beta)] = 0.62$ | | | | | | | | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 93.0% | 92.8% | 91.2% | 88.0% | 94.4% | 93.7% | 93.2% | 92.5% |
| Based on $\log[1 - R^2]$ | 93.8% | 93.9% | 92.6% | 92.1% | 94.8% | 94.1% | 93.9% | 93.5% |
| Based on $\log[R^2/(1 - R^2)]$ | 94.3% | 94.5% | 94.0% | 94.2% | 94.6% | 94.2% | 94.0% | 94.0% |
| Fisher's transformation | 91.4% | 91.4% | 90.4% | 90.0% | 92.4% | 91.9% | 91.4% | 90.8% |

0.9. Sample sizes and censoring were as in the case of a single covariate. The impact of varying degrees of correlation is shown in the tables.

The results show that the confidence intervals based on Taylor series approximations for $\log[R^2/(1 - R^2)]$ provide overall the most accurate coverage probabilities. This is true for all of the studied covariate distributions with a single variable (Tables I–III). The results are satisfactory even for censoring as high as 75%. The simple Fisher's transformation works well as a quick first approximation. Distribution of R^2 values for all simulations are provided in the Supporting Information. The limited study of correlation between the covariates is encouraging (Tables IV and V). The level of correlation appears to be well accounted for in the derived expressions. The Fisher transformation makes no use of the estimated

Table III. Simulation confidence intervals for R^2 with a covariate uniformly distributed based on Taylor series methods and Fisher's transformation.

| Method | $N = 100$ | | | | $N = 400$ | | | |
|--|-------------------------|-------|-------|-------|-----------|-------|-------|-------|
| | Percentage of censoring | | | | | | | |
| | 0% | 25% | 50% | 75% | 0% | 25% | 50% | 75% |
| $\beta = 1 \quad \mathcal{E}[R^2(\beta)] = 0.19$ | | | | | | | | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 92.0% | 90.5% | 90.3% | 86.4% | 94.9% | 94.3% | 93.7% | 93.1% |
| Based on $\log[1 - R^2]$ | 92.0% | 90.6% | 90.3% | 87.1% | 94.7% | 94.3% | 93.7% | 92.9% |
| Based on $\log[R^2/(1 - R^2)]$ | 96.5% | 96.7% | 96.6% | 95.1% | 96.2% | 95.6% | 95.9% | 97.1% |
| Fisher's transformation | 93.8% | 93.9% | 94.2% | 94.2% | 95.2% | 94.7% | 94.4% | 95.0% |
| $\beta = 2 \quad \mathcal{E}[R^2(\beta)] = 0.45$ | | | | | | | | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 93.6% | 93.6% | 92.3% | 90.1% | 95.2% | 95.2% | 95.2% | 94.7% |
| Based on $\log[1 - R^2]$ | 93.8% | 94.3% | 92.7% | 91.9% | 95.3% | 95.5% | 95.3% | 95.1% |
| Based on $\log[R^2/(1 - R^2)]$ | 95.0% | 95.5% | 95.6% | 96.1% | 95.7% | 95.8% | 96.0% | 96.1% |
| Fisher's transformation | 93.0% | 93.4% | 92.7% | 93.3% | 94.2% | 94.1% | 94.3% | 94.7% |
| $\beta = 3 \quad \mathcal{E}[R^2(\beta)] = 0.64$ | | | | | | | | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 92.6% | 93.0% | 91.2% | 90.0% | 94.2% | 94.7% | 95.0% | 93.6% |
| Based on $\log[1 - R^2]$ | 93.6% | 94.1% | 92.8% | 93.2% | 94.4% | 95.1% | 95.2% | 94.7% |
| Based on $\log[R^2/(1 - R^2)]$ | 94.0% | 94.9% | 93.8% | 95.1% | 94.4% | 95.2% | 95.4% | 95.1% |
| Fisher's transformation | 91.2% | 91.8% | 90.8% | 92.4% | 92.1% | 92.6% | 93.1% | 92.4% |

Table IV. Simulation confidence intervals for R^2 with two covariates normally distributed based on Taylor series methods and Fisher's transformation.

| Method | $N = 100$ | | | | | | | |
|--------------------------------|----------------------------------|-------|----------------------------------|-------|----------------------------------|-------|----------------------------------|-------|
| | $\rho = 0$ | | $\rho = 0.3$ | | $\rho = 0.6$ | | $\rho = 0.9$ | |
| | Percentage of censoring | | | | | | | |
| | 0% | 50% | 0% | 50% | 0% | 50% | 0% | 50% |
| $\beta_1 = \beta_2 = 0.5$ | | | | | | | | |
| | $\mathcal{E}[R^2(\beta)] = 0.13$ | | $\mathcal{E}[R^2(\beta)] = 0.16$ | | $\mathcal{E}[R^2(\beta)] = 0.18$ | | $\mathcal{E}[R^2(\beta)] = 0.20$ | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 90.9% | 89.5% | 92.3% | 90.1% | 92.9% | 90.4% | 92.3% | 91.4% |
| Based on $\log[1 - R^2]$ | 90.7% | 89.6% | 92.1% | 90.0% | 92.7% | 90.5% | 92.3% | 91.2% |
| Based on $\log[R^2/(1 - R^2)]$ | 96.2% | 96.3% | 96.8% | 96.6% | 96.7% | 96.4% | 96.6% | 96.7% |
| Fisher's transformation | 94.2% | 95.5% | 94.9% | 95.3% | 94.9% | 95.1% | 94.1% | 95.0% |
| $\beta_1 = \beta_2 = 1.0$ | | | | | | | | |
| | $\mathcal{E}[R^2(\beta)] = 0.32$ | | $\mathcal{E}[R^2(\beta)] = 0.36$ | | $\mathcal{E}[R^2(\beta)] = 0.41$ | | $\mathcal{E}[R^2(\beta)] = 0.44$ | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 93.7% | 92.7% | 94.1% | 93.1% | 94.0% | 92.8% | 93.8% | 91.9% |
| Based on $\log[1 - R^2]$ | 93.9% | 93.1% | 94.6% | 93.4% | 94.4% | 93.8% | 94.1% | 93.4% |
| Based on $\log[R^2/(1 - R^2)]$ | 96.2% | 96.7% | 95.5% | 96.7% | 95.7% | 96.1% | 95.3% | 95.7% |
| Fisher's transformation | 94.1% | 94.1% | 94.0% | 94.3% | 94.9% | 94.0% | 93.4% | 92.8% |
| $\beta_1 = \beta_2 = 1.5$ | | | | | | | | |
| | $\mathcal{E}[R^2(\beta)] = 0.48$ | | $\mathcal{E}[R^2(\beta)] = 0.54$ | | $\mathcal{E}[R^2(\beta)] = 0.58$ | | $\mathcal{E}[R^2(\beta)] = 0.62$ | |
| Taylor series methods | | | | | | | | |
| Based on R^2 | 93.6% | 92.9% | 93.5% | 93.1% | 93.3% | 92.1% | 93.2% | 91.5% |
| Based on $\log[1 - R^2]$ | 94.0% | 94.2% | 94.0% | 94.1% | 94.5% | 93.3% | 94.2% | 93.1% |
| Based on $\log[R^2/(1 - R^2)]$ | 95.0% | 95.7% | 95.0% | 95.7% | 94.9% | 94.5% | 94.5% | 94.1% |
| Fisher's transformation | 92.4% | 93.5% | 92.2% | 93.4% | 92.2% | 92.0% | 91.8% | 91.2% |

correlations and, although it is less impressive than the Taylor series approximations, it still provides a good working guide for practical use. Here, these cases have been limited to bivariate normal, and it may be worth extending this study to deal with binary outcomes. It is less clear cut however how to carry out such studies because many different copula-type models could be suggested.

| Table V. Simulation confidence intervals for R^2 with two covariates normally distributed based on Taylor series methods and Fisher's transformation. | | | | | | | | |
|---|----------------------------------|-------|-----------------------------------|-------|-----------------------------------|-------|----------------------------------|-------|
| $N = 400$ | | | | | | | | |
| Method | $\rho = 0$ | | $\rho = 0.3$ $\rho = 0.6$ | | | | $\rho = 0.9$ | |
| | 0% | 50% | Percentage of censoring | | Percentage of censoring | | 0% | 50% |
| | | | 0% | 50% | 0% | 50% | 0% | 50% |
| $\beta_1 = \beta_2 = 0.5$ | | | | | | | | |
| Taylor series methods | $\mathcal{E}[R^2(\beta)] = 0.11$ | | $\mathcal{E}[R^2(\beta)] = 0.14$ | | $\mathcal{E}[R^2(\beta)] = 0.16$ | | $\mathcal{E}[R^2(\beta)] = 0.18$ | |
| Based on R^2 | 93.6% | 94.0% | 95.1% | 93.4% | 95.0% | 94.2% | 94.4% | 94.9% |
| Based on $\log[1 - R^2]$ | 93.6% | 93.8% | 95.0% | 93.4% | 94.9% | 94.1% | 94.4% | 94.6% |
| Based on $\log[R^2/(1 - R^2)]$ | 96.2% | 96.7% | 96.5% | 96.4% | 96.2% | 97.2% | 95.8% | 96.7% |
| Fisher's transformation | 94.5% | 95.8% | 95.6% | 94.9% | 95.2% | 95.8% | 94.8% | 95.7% |
| $\beta_1 = \beta_2 = 1.0$ | | | | | | | | |
| Taylor series methods | $\mathcal{E}[R^2(\beta)] = 0.29$ | | $\mathcal{E}[R^2(\beta)] = 0.34$ | | $\mathcal{E}[R^2(\beta)] = 0.38$ | | $\mathcal{E}[R^2(\beta)] = 0.42$ | |
| Based on R^2 | 95.1% | 95.0% | 94.8% | 95.4% | 94.9% | 95.3% | 94.5% | 95.4% |
| Based on $\log[1 - R^2]$ | 95.0% | 95.2% | 95.1% | 95.4% | 94.8% | 95.3% | 94.8% | 95.7% |
| Based on $\log[R^2/(1 - R^2)]$ | 95.7% | 96.5% | 95.3% | 96.1% | 95.3% | 96.0% | 94.8% | 96.0% |
| Fisher's transformation | 94.5% | 95.2% | 94.1% | 94.8% | 93.8% | 94.6% | 93.3% | 94.8% |
| $\beta_1 = \beta_2 = 1.5$ | | | | | | | | |
| Taylor series methods | $\mathcal{E}[R^2(\beta)] = 0.46$ | | $\mathcal{E}[R^2(\beta)] = 0.524$ | | $\mathcal{E}[R^2(\beta)] = 0.568$ | | $\mathcal{E}[R^2(\beta)] = 0.60$ | |
| Based on R^2 | 95.4% | 95.5% | 95.6% | 95.1% | 95.1% | 94.9% | 94.8% | 94.7% |
| Based on $\log[1 - R^2]$ | 95.4% | 95.5% | 95.7% | 95.4% | 95.5% | 95.0% | 95.0% | 95.3% |
| Based on $\log[R^2/(1 - R^2)]$ | 95.6% | 96.0% | 95.9% | 95.7% | 95.4% | 95.5% | 95.2% | 95.4% |
| Fisher's transformation | 94.0% | 94.3% | 93.9% | 94.0% | 93.2% | 92.8% | 92.5% | 93.1% |

5. Examples

Our first application is taken from [1] as we can easily calculate confidence intervals based on the Fisher's transformation approach knowing the number of observed events and the measure estimate. Our results are in good agreement with those obtained using bootstrap method except for values larger than 0.9 (R_{PM}^2 , R_{OF}^2 and R_{XO}^2 for the Leg ulcer study) where the Fisher's transformation provided narrower confidence intervals and for values smaller than 0.25 where it provided larger confidence intervals (Table VI).

The Breast Cancer data involved 686 individuals with complete information of which 299 experienced the event of interest [10]. The final model suggested by [11] included age, tumour grade, number of positive lymph nodes, progesterone receptor and hormonal therapy [1]. Data were downloaded from <http://www.homepages.ucl.ac.uk/ucakjpr/>. The lymphoma data come from a study on diffuse large B cell lymphoma. The authors in [12] evaluated the extent to which the continuous Rosenwald gene score adds to the International Prognostic Index (IPI) in the prediction of survival in the 73 patients of the validation set for which the IPI values were available. Data were downloaded from <http://lmpp.nih.gov/DLBCL/>. For both studies, confidence intervals obtained using the Taylor approximations are in good agreement with those obtained via the Fisher's transformation or those obtained by bootstrap methods (Table VII).

Our last example is taken from [13] and is a set of 40 survival times for lung cancer patients, with seven covariates for each patient. The main purpose of the study was to compare the effects of two chemotherapy treatments in prolonging survival time, while assessing the effects of other factors such as age and months under treatment. Various analyses by [13] indicate that performance status is highly significant, and cell type is marginally significant; in particular, cell-type adeno is associated with an increased risk of failure. We then restrict attention to these two covariates having significant effects. Results are shown in Table VIII.

6. Discussion

Interest in predictive measures for survival models has grown in recent years. One reason for this comes from the need to quantify the impact on outcome of several, possibly a large number, of biomarkers or genetic information. The impact of markers can be subtle so that finding relevant effects leads us to careful

Table VI. Estimates of explained variation measures and 95% confidence intervals based on Fisher's transformation method and bootstrapping.

| | Leg ulcer study | | Breast cancer study | | Lymphoma study | |
|--------------|------------------|-------------|---------------------|-------------|------------------|-------------|
| N events | 97 | | 299 | | 48 | |
| N subjects | 200 | | 686 | | 73 | |
| Measures | Fisher | Bootstrap | Fisher | Bootstrap | Fisher | Bootstrap |
| R_{PM}^2 | 0.94 (0.91–0.96) | (0.80–0.98) | 0.27 (0.19–0.36) | (0.21–0.35) | 0.23 (0.05–0.45) | (0.11–0.42) |
| R_{OF}^2 | 0.92 (0.88–0.95) | (0.80–0.98) | 0.37 (0.29–0.46) | (0.30–0.46) | 0.32 (0.11–0.54) | (0.14–0.59) |
| R_{XO}^2 | 0.93 (0.90–0.95) | (0.81–0.98) | 0.38 (0.29–0.47) | (0.30–0.48) | 0.27 (0.08–0.49) | (0.08–0.54) |
| R_D^2 | 0.52 (0.37–0.65) | (0.44–0.65) | 0.28 (0.20–0.37) | (0.21–0.35) | 0.23 (0.05–0.45) | (0.11–0.40) |
| R_R^2 | 0.58 (0.44–0.70) | (0.46–0.71) | 0.29 (0.20–0.38) | (0.23–0.38) | 0.21 (0.04–0.43) | (0.10–0.42) |

Measures and bootstrap confidence intervals estimates obtained from [1]. The R_{OF}^2 of Choodari-Oskooei *et al* (2012a) corresponds to our $R^2(\hat{\beta})$.

Table VII. Estimates of $R^2(\hat{\beta})$ and 95% confidence intervals based on Taylor series methods and Fisher's transformation method for breast cancer and lymphoma data.

| | Breast cancer study | Lymphoma study |
|--------------------------------|------------------------------|------------------------------|
| $R^2(\hat{\beta})$ | 0.37 | 0.32 |
| Taylor series methods | 95% CI of $R^2(\hat{\beta})$ | 95% CI of $R^2(\hat{\beta})$ |
| Based on R^2 | (0.28–0.47) | (0.12–0.53) |
| Based on $\log[1 - R^2]$ | (0.27–0.46) | (0.09–0.50) |
| Based on $\log[R^2/(1 - R^2)]$ | (0.29–0.47) | (0.16–0.55) |
| Fisher's transformation | (0.28–0.46) | (0.12–0.54) |

Table VIII. Estimates of $R^2(\hat{\beta})$ and 95% confidence intervals (CI) based on Taylor series methods and Fisher's transformation method for lung cancer data.

| Variable | $\hat{\beta}_i$ | $R^2(\hat{\beta}_1, \hat{\beta}_2)$ | CI method | Lung cancer study 95% CI of $R^2(\hat{\beta}_1, \hat{\beta}_2)$ |
|--------------------|-----------------|-------------------------------------|--------------------------------|--|
| Performance status | -0.058 | 0.58 | Based on R^2 | (0.35–0.81) |
| Adeno | 1.070 | | Based on $\log(1 - R^2)$ | (0.27–0.76) |
| | | | Based on $\log(R^2/(1 - R^2))$ | (0.35–0.78) |
| | | | Fisher's transformation | (0.34–0.76) |

consideration of both the fit and the predictive power of the models we use. Before dealing with the more subtle effects of biomarkers, it is important to correctly account for the prognostic role played by the more usual, often well-known, risk factors such as stage or grade of disease, age and other clinical factors that have been well studied. The effect of biomarkers can easily be masked or lost to uncertainty when we include the main clinical or demographic factors in any model, and so it is important to have some idea as to the precision of our predictive measures, given the main variables, before looking for more subtle effects. A very accurate assessment of this precision is not really necessary, but ballpark figures can indicate whether or not we have a good chance of picking up weaker effects. Goodness of fit plays a part here too. We have not considered that aspect in this work. Our results show quite good agreement with those obtained by Choodari-Oskooei [1]. An interval based on the Fisher transformation will be helpful in general situations. A simple application of the inverse hyperbolic tangent function, along with the added amount of predictive effect that we would like to be able to detect, leads to a direct assessment of the number of failures that we will need to observe. That can be of use in a sample size calculation or, otherwise, as an indicator of what we can reasonably hope to accomplish given the constraints of a particular study.

We find broad agreement with intervals obtained via bootstrap resampling. It might be argued that the effectiveness of the bootstrap obviates the need for more involved calculations. While there is something to be said for that viewpoint, we still think it is useful to have analytical expressions. The one based on the Fisher transform appears to perform remarkably well given its simplicity and it is simple enough to be

quickly evaluated by hand, or even mentally, in the context where computer-based resampling may not be readily available. Some reassurance can be obtained from the fact that there is broad agreement between all of the approaches. Nominal coverage of the confidence intervals cannot be described as highly accurate but is nonetheless good and exhibits only a very weak dependence, if any, on the strength of association itself, the degree of censoring and the amount of correlation between the variables, at least in the limited cases studied. Improvements are very likely possible but because our purpose is one of obtaining a rough ballpark guide rather than one of high accuracy, the extra investment in calculation may not be justified.

Acknowledgements

The authors would like to thank the editors and reviewers for a detailed review of an earlier version of this work. In particular, we would like to thank them for having identified an important error in the development.

References

1. Choodari-Oskooei B, Royston P, Parmar MKA. Simulation study of predictive ability measures in a survival model I: explained variation measures. *Statistics in Medicine* 2012; **31**:2627–2643.
2. Choodari-Oskooei B, Royston P, Parmar MK. A simulation study of predictive ability measures in a survival model II: explained randomness and predictive accuracy. *Statistics in Medicine* 2012; **31**:2644–2659.
3. O'Quigley J, Flandre P. The predictive capability of proportional hazards regression. *Proceedings of the National Academy of Sciences of the United States of America* 1994; **91**:2310–2314.
4. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* 1972; **34**:187–220.
5. O'Quigley J, Xu R. Explained variation in proportional hazards regression. In *Handbook of Statistics in Clinical Oncology*, Crowley J (ed.) Marcel Dekker: New York, 2001; 397–409.
6. Schoenfeld DA. Partial residuals for the proportional hazards regression model. *Biometrika* 1982; **69**:239–241.
7. Andersen P K, Gill R D. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 1982; **10**:1110–1120.
8. Deutsch R. *Survival Prediction Following HIV Infection: Interval Censored Infection Times and Subsequent Cognitive Impairment and Mortality*. University of California: San Diego, 1996. 9714837.
9. Fisher R. On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron* 1921; **1**:3–32.
10. Schumacher M, Bastert G, Boja H, Hübner K, Olschewski M, Sauerbrei W, Schmoor C, Beyerle C, Neumann RL, Rauschecker HF. Randomized 2×2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. German Breast Cancer Study Group. *Journal of Clinical Oncology* 1994; **12**:2086–2093.
11. Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society (Series A)* 2002; **162**:71–94.
12. Dunkler D, Michiels S, Schemper M. Gene expression profiling: does it add predictive accuracy to clinical characteristics in cancer prognosis? *European Journal of Cancer* 2007; **43**:745–751.
13. Lawless JF. *Statistical Models and Methods for Lifetime Data*. Wiley: New York, 1982.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article.