HHS Public Access

Author manuscript

Commun Stat Simul Comput. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

Commun Stat Simul Comput. 2018 ; 47(2): 380–391. doi:10.1080/03610918.2016.1177072.

A Study of R^2 Measure under the Accelerated Failure Time Models

Priscilla H Chan[1], **Ronghui Xu**[2,3,*], and **Christina D Chambers**[1,2]

[1]Department of Pediatrics, University of California, San Diego

[2]Department of Family and Preventative Medicine, University of California, San Diego

[3]Department of Mathematics, University of California, San Diego

Abstract

For right-censored data the accelerated failure time (AFT) model is an alternative to the commonly used proportional hazards regression model. It is a linear model for the (log-transformed) outcome of interest, and is particularly useful for censored outcomes that are not time-to-event, such as laboratory measurements. We provide a general and easily computable definition of the R^2 measure of explained variation under the AFT model for right-censored data. We study its behavior under different censoring scenarios and under different error distributions; in particular, we also study its robustness when the parametric error distribution is misspecified. Based on Monte Carlo investigation results, we recommend the log-normal distribution as a robust error distribution to be used in practice for the parametric AFT model when the R^2 measure is of interest. We apply our methodology to an alcohol consumption during pregnancy data set from Ukraine.

Keywords

censoring type; error distribution; explained variation; log-normal distribution; semiparametric AFT model; transformation models

1 Introduction

The R^2 measure of explained variation for right-censored data has been studied since early days of survival analysis up until very recently. Much of the work has been done under the popular proportional hazards regression model. An early reference is Harrell (1986), and a comprehensive review and comparison can be found in Schemper and Stare (1996) for works published up until then. O'Quigley and Xu (2012) also thoroughly examined the concepts of explained variation as well as explained randomness under the proportional hazards model. Very roughly speaking, there are three general ways to define such a measure: squared correlation coefficient; residual based; and information based. All three as well as the R^2 measure itself is best understood under the linear regression model without

*Corresponding author: 9500 Gilman Drive, MC 0112, La Jolla, CA 92093-0112; rxu@ucsd.edu.

censoring. In addition, the definitions are generally model dependent, i.e. specifically for the proportional hazards model.

The work of this paper was motivated by a prospective cohort study conducted among pregnant women in Ukraine as a part of the Collaborative Initiative on Fetal Alcohol Spectrum Disorders (Arenson et al., 2010; Mattson et al., 2010, CIFASD). One of the goals of the study was to identify predictors of maternal alcohol consumption in pregnancy. In this study, all women who came in for a routine prenatal visit were screened with questions on alcohol consumption, drug or vitamin exposures, demographics and pregnancy history. During the interview, a woman reported the number of days in a month that she consumed 1 to 2, 3 to 4, or 5 or more drinks on an occasion around conception as well as in the most recent month. To capture the frequency and amount of alcohol consumed, we decided to examined the average absolute ounces of alcohol consumed by a woman per day and per drinking day. Such calculations created right-censoring data, because for those who reported 5 or more drinks on an occasion, the exact amount was unknown. If we ignore it and fit it with ordinary least squares regression, which could not properly handle right-censored outcome data, the R^2 in the naive fit would show bias with increasing amount of censoring, as illustrated in Figure 1.

To handle this type of right-censored outcome data, the popular Cox proportional hazard regression model does not provide a natural interpretation as it models the hazard function, while our outcome is not time-to-event. Instead, we consider the accelerated failure time (AFT) model. The AFT model is a linear regression model, albeit typically with logarithm transformation on the outcome, and predicts the mean amount of drinking given the other maternal characteristic variables. As in many practical problems, we would like to know the ability of the maternal characteristics in predicting maternal drinking as reflected by an R^2 measure. While the R^2 measure is well-known under the linear models with no censoring in the data, to our best knowledge no R^2 measures have been proposed in the literature under the AFT model with right-censored data. As we will see, however, it is not difficult to develop such a measure.

In the following we first define an R^2 measure under the AFT model. We then study its performances via simulation in Section 3, including when the model is missepecified. In Section 4 we apply the measure to the alcohol consumption during pregnancy data set from Ukraine. We conclude with discussion in Section 5.

2 R^2 measure under the AFT model

The accelerated failure time model can be written

$$Y = Z^{\top}\beta + \sigma\varepsilon, \quad (1)$$

where Y is the response variable, often the logarithm of the uncensored outcome T of interest, Z is a vector of covariates or independent variables, β is a vector of unknown regression parameters, σ is an unknown scale parameter, and ε is an error term with fixed

variance. While the semiparametric AFT model has been studied in the statistical literature where the distribution of ε is unspecified, the parametric AFT model is more widely used in practice due to its ease of computation and availability in common statistical software, and it is our focus here. However, the definition below can be easily extended to the semiparametric AFT model, as will be discussed later. The commonly used parametric distributions for ε are: exponential, logistic, log-logistic, log-normal, Gaussian, and Weibull.

To define an R^2 measure under model (1), we recall that under the linear regression model with no censoring, it can be defined as one minus the ratio of residual sum of squares to the total sum of squares. With right-censored data, however, the sums of squares are not straightforward to obtain. Instead, we notice that R^2 should reflect the proportion of explained variation over the total variation, and the explained variation can be obtained by subtracting the residual variation from the total variation. That is, we wish to estimate

$$\Omega^2 = 1 - \frac{\mathrm{Var}(\sigma\varepsilon)}{\mathrm{Var}(Y)}. \quad (2)$$

Since ε is independent of Z, from (1) we have $\mathrm{Var}(Y) = \mathrm{Var}(Z^\top\beta) + \sigma^2\mathrm{Var}(\varepsilon)$. For a sample with $i = 1, ..., n$ subjects, after estimating β and σ under the AFT model (1), we define

$$R^2 = 1 - \frac{\hat{\sigma}^2\mathrm{Var}(\varepsilon)}{\widehat{\mathrm{Var}}(Z^\top\beta) + \hat{\sigma}^2\mathrm{Var}(\varepsilon)}, \quad (3)$$

where $\widehat{\mathrm{Var}}(Z^\top\beta) = \sum_{i=1}^{n}\{Z_i^\top\hat{\beta} - (\sum_{j=1}^{n} Z_j^\top\hat{\beta})/n\}^2/(n-1)$ is the sample variance of the $Z_i^\top\hat{\beta}$'s, and $\mathrm{Var}(\varepsilon)$ is known under the model. For the six commonly used distributions mentioned above, $\mathrm{Var}(\varepsilon) = \pi^2/3$ under the logistic and log-logistic models, $\pi^2/6$ under the Weibull and exponential models, and 1 under the Gaussian and log-normal models. Since the above sample variance consistently estimates $\mathrm{Var}(Z^\top\beta)$, and the parameters are consistently estimated under model (1), we know that the R^2 defined in (3) consistently estimates Ω^2 defined in (2).

3 Simulation studies

While the computation of R^2 defined in (3) is straightforward, in practice one needs to choose an error distribution before fitting the AFT model (1). In addition, we would also like to know how censoring and covariate distribution might affect the R^2 in finite samples.

3.1 Simulation setup

We simulated data sets under model (1). Data were generated with the intercept $\beta_0 = 0$ without loss of generality. We considered four types of right-censoring scenarios: no censoring, type I, type II, and random censoring. While random censoring commonly occurs in clinical studies, type I and type II censorings are often the case for laboratory or engineer

studies, for which the AFT model might often be used. For each of the censoring scenarios except no censoring, we generated q = 25%, 50% and 75% censoring. For type I censoring, the censoring time C was fixed, which was the 100 × (1 − q)-th percentile of the distribution of Y; for type II censoring, the censoring time C was the 100 × (1 − q)-th percentile of the sample of Y_i's; for random censoring we used Uniform (0, τ) distribution, where τ was chosen so that a pre-specified censoring percentage was achieved. We used R function survreg() written by Therneau and Lumley to fit the AFT model. For each combination of the parameter values, results from 2000 simulations, each with sample size 100, were given below unless otherwise specified. Other sample sizes have also been considered, but due to the limitation on space we show the most representative results here.

We considered four different covariate distributions for Z under model (1): Bernoulli, Normal, Uniform, or Exponential. These included discrete, continuous, symmetric, and skewed distributions. All of the distributions had Var(Z) = 0.25. Note that it is necessary to fix Var(Z) so that the R^2 values can be compared for a given value of β; otherwise a scale change in Z is simply equivalent to multiplying β by a non-zero constant. We set β values from 0.5 to 4 with 0.5 increments.

For error distribution we generated ε according to logistic, log-logistic, Weibull, Gaussian and log-normal, with $\sigma^2 = 3/\pi^2$, $3/\pi^2$, $6/\pi^2$, 1 and 1 respectively, so that all Var($\sigma\ \varepsilon$) = 1. For exponential error distribution, $\sigma = 1$ by definition, so that Var($\sigma\ \varepsilon$) = $\pi^2/6$.

Finally we generated data with Z distributed as bivariate normal with variances 0.25 and correlation coefficient ρ = 0, 0.2 or 0.5, β_1 and β_2 varied from −2 to 2 with 0.5 increment, and $\varepsilon \sim \mathcal{N}(0, 1)$.

3.2 Simulation results

Effects of censoring, covariate and error distributions—In Table 1 in order to compare the R^2 values versus Ω^2 across different error distributions, we set $\Omega^2 = 0.5$ by construction. This is done by setting $\beta = 2$ under all except exponential error distribution, where we set $\beta = \sqrt{2/3}\pi$, together with the σ values given in the above setup. As shown in the table, the R^2 measure shows desirable performance when the error distribution is specified correctly. The R^2 measure has the least variation with the exponentially distributed error, most likely because $\sigma = 1$ is known. Figure 1 shows the naive R^2 values when the right-censored outcome data was fitted with ordinal least-squares ignoring the censoring indicator. It used a large sample size of 10000, $\varepsilon \sim \mathcal{N}(0, 1)$ and $Z \sim \mathcal{N}(0, 0.25)$. It is clear that censoring can have a profound effect on the R^2 values if a proper method is not used to analyze the right-censored data.

Figure 2 demonstrates the behavior of R^2 for different covariate distributions. Here $\varepsilon \sim \mathcal{N}(0, 1)$; the results under other true error distributions are similar. As shown in the figure the R^2 values are basically unaffected by the underlying distribution of Z (except binary with 75% censoring; see below). The error bars represent the standard deviation from the simulations. We see that the R^2 has slightly more variation when Z is exponentially distributed, and the least variation when Z is Normal or Uniform.

The effect of different censor scenarios is also illustrated in Figure 2. Generally, the R^2 measure behaves very similarly under different types of censoring. In addition, it has less variation with decreasing amount of censoring and increasing sample size as expected (data not shown). For 75% censoring and when Z is Bernoulli, over 95% of those subjects with $Z = 1$ are censored when $\beta > 2$, and $\hat{\beta}$ is overestimated. This has resulted in usually large R^2 values as shown in Figure 2.

Misspecified error distribution—Figure 3 shows for each of the six error distributions that are used to generate the data, the R^2 values when each of these six distributions are fitted to the data under model (1), respectively. Note that when the true error is Gaussian or logistic, only these two models are used to fit the data since the outcome can be potentially negative. In Figure 3 the parameter values are chosen so that the true explained variation $\Omega^2 = 0.5$. We see that among all the fitted error distributions, log-normal (in red color) is the most robust against distributional misspecification. Exponential and Weibull models underestimate the R^2 in all cases except when they are the true error distribution. Closer examination of the parameter estimates under the misspecified error distributions reveals that both $\hat{\beta}$ and $\hat{\sigma}$ can deviate from their true values when the model is wrong; interestingly when the log-normal error distribution is used to fit the data, the deviations in the parameter estimates appear to 'cancel out' to give nearly correct R^2 values as compared to Ω^2.

In Figure 4 we consider different values of Ω^2 from 0.1 to 0.8, with increments of 0.1. We used normally distributed Z with 25% type I censoring; results for other types of censoring are qualitatively comparable (data not shown). The estimated values are plotted near the grid points for the corresponding Ω^2 values, but are slightly shifted left or right so that different colors can be as visible as possible. The color scheme is the same as in Figure 3 for the fitted error distributions. For each column of the true error distribution, we should compare to the results when the true distribution is used to fit the model; for example, in the third column when the true distribution is log-normal, we should compare all the other colors to red, which represents when the true log-normal model is fitted (hence the results are reliable). Across the top row we see that overall log-normal and log-logistic give R^2 values the closest to when the true error distribution is fitted (remember to only compare the vertical values near the grid point, and not the horizontal distrances). Upon closer examination using plots like in Figure 3 (data not shown) we see that at the lower end of Ω^2 values like 0.1, log-logistic does slightly better than log-normal, in terms of the robustness under consideration. But overall, log-normal appears to perform the best, with log-logistic a close second. The middle and bottom rows show the $\widehat{\mathrm{Var}}(Z^\top\beta)$ and $\hat{\sigma}$ values, respectively. And we further discuss the goodness-of-fit issue in the last section.

Multiple covariates—Figure 5 shows the R^2 values (as %) with varying correlation coefficients for the two covariate cases with 0% and 25% random censoring. R^2 measure also behaves similarly with type I and II censoring (data not shown).

4 Drinking during pregnancy data

We now return to the CIFASD alcohol consumption during pregnancy study. As mentioned earlier the purpose of this study was to evaluate the patterns of alcohol consumption in

pregnancy as reported by women in Ukraine and to describe maternal predictors of potentially harmful alcohol consumption to help inform future intervention and prevention strategies. The study screened 11,909 pregnant women from two study sites in Ukraine between 2007 and 2012, and 10,976 of them were classified as 'ever-drinking'. These 'ever-drinking' women were asked to recall the number of days they had 1 or 2 drinks, 3 or 4 drinks, or 5 or more drinks during the relevant time period. In the calculation 1 or 2 drinks was coded as 1.5 drinks, and 3 or 4 drinks was coded as 3.5 drinks. The category of 5 or more drinks created right censoring in the summary measure below, as the actual number of drinks was unknown. Alcohol consumption in response to these questions was summarized as the average number of drinks per day over the month for which the woman was reporting as reflection of the overall quantity of alcohol consumed (around conception, and during the most recent month). More specifically, this is equal to $(1.5x + 3.5y + 5z)/30$ where x, y and z are the number of days they had 1 or 2 drinks, 3 or 4 drinks, or 5 or more drinks, respectively. Overall 7262 women reported having > 0 drinks around conception, and 5841 women reported having > 0 drinks in the most recent month. Our analyses below are restricted to these subsets of women. There were totally about 7% right-censored observations.

In additional to the quantity and frequency of alcohol consumption, women were also asked to respond to seven standard screening questions for risky drinking, including binary predictors for inability to remember activities after a drinking episode (Amnesia), feeling annoyed by other's criticism of the respondent's drinking (Annoy), the desire to cut down on drinking (Cut-Down), the need for an "eye-opener" (Eye-opener), guilt regarding drinking (Guilty), others are concerned about the respondent's drinking (Worry), and ability to hold six or more drinks (Tolerance). The number of drinks a woman can hold, i.e. before passing out, falling asleep, or becoming too sick to continue, was also recorded (HOLD). The seven questions above comprise the CAGE (Cut-down + Annoy + Guilty + Eye-opener), the TWEAK (2×Tolerance + 2×Worry + Eye-opener + Amnesia + Cut-down), and the T-ACE(2×Tolerance + Annoy + Cut-down + Eye-opener) scores, that have five levels (0 – 4), eight levels (0 – 7), and six levels (0 – 5), respectively (Russell et al., 1996).

We fit the AFT model (1) to each of the two alcohol consumption outcome measures (around conception, and during the most recent month), with the risky drinking predictors described above, as well as relevant maternal characteristics variables. The maternal characteristics variables were selected from twenty-two covariates using stepwise procedures (Chambers et al., 2014), and included smoking status, age when first started to drink, marital status, education, gravidity, etc. The prediction models were built with each of the risky drinking predictors alone, and with each of the CAGE, TWEAK and T-ACE scores plus the selected maternal characteristics covariates. The R^2 values were computed for each model, and Figure 6 shows these values under various error distributions for drinking around conception; the results for the most recent month in pregnancy are qualitatively similar, but with slightly higher R^2 values (see below). Logistic error distribution gave notably higher R^2 values for some models, when compared to the other error distributions. Closer examination of the model fits under logistic revealed that the scale parameter σ as well as the error variance was estimated to be very small in magnitude. The same was true to a degree for Gaussian error, leading to also relatively large R^2 values. While in this case we do not know

what the true model is, in simulation studies we have also noticed relatively poorly estimated scale parameters when these two error distributions are used (Figure 4).

As the log-normal error distribution was found to be generally a robust choice according to our simulation results above, we also tabulated the R^2 values under log-normal in Table 2, for both around conception and during the most recent month drinking. We can see that the single predictor HOLD gives the highest R^2 values for both outcomes, explaining 41% of variation in drinks per day during the most recent month in pregnancy. The dichotomized version of it, Torelance, also explains substantial amounts of variation, and is included in the calculation of TWEAK and T-ACE. Overall, the predictors are better at explaining the variation in the amount of drinking per day in the most recent month than around conception.

5 Discussion

In this paper we have studied an R^2 measure under the AFT model that is easy to calculate. A similar approach was suggested in Heller (2012) for the proportional hazards model, but on the risk scale instead of the variance scale as we consider here. The decomposition of the variance of the outcome variable was also used in Xu (2003) for an R^2 measure under the linear mixed effects models. The measure is robust with respect to the covariate distribution, censoring, and the true underlying error distribution. In practice the error distribution can be misspecified under the parametric AFT model, and through simulation studies we have found the log-normal error distribution to be the relatively robust against such misspecification. Here we emphasize that the R^2 measure, while related to the concept of goodness-of-fit, is really a measure of explained variation. This can be easily seen in a simple linear regression example where the regression slope is very close to zero but with no violation of the model assumptions, in this case the R^2 is very close to zero albeit the fit is good. In our opinion, the R^2 measure is designed to capture the ability of the covariates in predicting the response; while this is often model dependent, it is desirable to be robust against model misspecifications. That is, the R^2 values ideally should not change substantially even if some of the model assumptions are violated. It is in this light that we recommend the use of the log-normal error distribution under the parametric AFT model in conjunction with the use of the R^2 measure defined in this paper.

In applications of any statistical model ideally we would like to be able to assess the goodness-of-fit and, while the R^2 is not a measure of fit, improved fit could lead to greater values of R^2. Unfortunately with right-censored data, although attempts have been made in the literature to check the parametric distributional assumptions, they tend not to be very sensitive to model departures and their ‘practical utility is limited’; see O’Quigley and Xu (1998) and references therein. More successes have been achieved under the semiparametric settings such as checking the proportional hazards assumption (O’Quigley and Xu, 1998; O’Quigley, 2003). These facts highlight the need to develop more effective model diagnostic tools under the parametric models for right-censored data and, perhaps until then, the importance of robustness of the approaches utilized to analyze data including the R^2 measure.

It is known that for multivariate linear regression, adding new regressors, even when unrelated to the outcome will increase the R^2 value. This is directly due to the least-squares estimation method, which is equivalent to the maximum likelihood estimation (MLE) under the normally distributed errors. For other error distributions under the AFT model, this may not be exactly the case for the MLE, but empirically we do observe that adding covariates tend to increase the R^2 value. Some statisticians use the so-called adjusted R^2. Our approach in practice has been to bare in mind this property of the R^2, and see if the increase in R^2 is 'worth' the addition of the extra covariate(s). In our example of the last section, adding the covariates to TWEAK increased the R^2 for the around conception outcome from about 0.20 to 0.25, which might be considered a nontrivial increase. Alternatively, one could also adopt resampling methods to estimate the potential bias. Bootstrap (bias-corrected) confidence intervals for Ω^2 were considered in Xu (1996).

We have defined an R^2 measure that directly handles multiple covariates. Partial coefficients can also be defined. For example if there are two sets of covariates Z_1 and Z_2, one may ask the question of how much added value is Z_2 in explaining the outcome after Z_1 has been accounted for. Denote R^2_1 the measure under the model using only Z_1, and R^2_{1+2} the measure under the model using both Z_1 and Z_2. Then one can define the partial coefficient $R^2_{2|1} = 1 - (1 - R^2_{1+2})/(1 - R^2_1)$. Using the TWEAK ($Z_1$) plus covariates ($Z_2$) example above, $R^2_{2|1} = 1 - (1 - 0.254)/(1 - 0.205) = 0.06$ in that case. An alternative approach, however, is to define coefficients for the model with a single regressor, and to define partial coefficients, and then to build multivariate coefficients using a formula similar to the above. Such approaches were considered in details in O'Quigley and Flandre (1994). These different approaches lead to the same multivariate coefficient in linear regression, but is not necessarily the case under the AFT model here, which depends on how one constructs the simple and partial coefficients.

While we have focused on the parametric AFT model which are widely used in practice, formulas (2) and (3) are readily extended to the semiparametric transformation models, which includes the Cox proportional hazards model as a special case. It is clear that as long as the regression coefficients and the error variance are consistently estimated, the R^2 measure consistently estimates Ω^2, which has a clear interpretation as explained variation. In the case of transformation models, this is the explained variation in the transformed response variable, which is most likely different from the explained variation in the response variable in its original scale. However, in the context of the transformation model being used to analyze the data, it is perhaps this explained variation in the transformed response that is the more relevant.

Another related approach that has been recently advocated in the literature is information theoretical, also called explained randomness (Kent, 1983; Kent and O'Quigley, 1988; O'Quigley et al., 2005; Preseley et al., 2011). This approach is also easy to calculate as it is typically based on the likelihood ratio statistic that is often part of the output from a model fitting software. It is possible, however, that this likelihood based approach might be more sensitive to distributional misspecifications. Further investigation is needed in order to verify

whether it can be as robust as the R^2 measure defined in (3) together with the log-normal error distribution under the AFT model with right-censored data.

Acknowledgments

The data used in this paper was collected with the support of NIH Research Grant No. U01AA014835.

References

Arenson AD, Bakhireva LN, Chambers CD, Deximo CA, Foround T, Jacobson JL, Jacobson SW, Jones KL, Mattson SN, May PA, Moore ES, Ogle K, Riley EP, Robinson LK, Rogers J, Streissguth AP, Tavares MC, Urbanski J, Yezerets Y, Surya R, Stewart CA, Barnett WK. Implementation of a shared data repository and common data dictionary for fetal alcohol spectrum disorders research. Alcohol. 2010; 44:643–647. [PubMed: 20036486]

Chambers CD, Yevtushok L, Zymak-Zakutnya N, Korzhynskyy Y, Ostapchuk L, Akhmedzhanova D, Chan PH, Xu R, Wertelecki W. Prevalence and predictors of maternal alcohol consumption in 2 regions of ukraine. Alcoholism: Clinical and Experimental Research. 2014; 38:1012–1019.

Harrell FE. The PHGLM procedure. SAS Supplement Library User's Guide, Version 5. 1986:437–466.

Heller G. A measure of explained risk in the proportional hazards model. Biostatistics. 2012; 13:315–325. [PubMed: 22190711]

Kent JT. Information gain and a general measure of correlation. Biometrika. 1983; 70:163–174.

Kent JT, O'Quigley J. Measures of dependence for censored survival data. Biometrika. 1988; 75:525–534.

Mattson SN, Foround T, Sowell ER, Jones KL, Coles CD, Fagerlund A, Riley EP. the CIFASD Group. Collaborative initiative on fetal alcohol spectrum disorders: methodology of clinical projects. Alcohol. 2010; 44:635–641. [PubMed: 20036488]

O'Quigley J. Khmaladze-type graphical evaluation of the proportional hazards assumption. Biometrika. 2003; 90:577–584.

O'Quigley J, Flandre P. Predictive capability of proportional hazards regression. Proc of the National Academy of Science USA. 1994; 91:2310–2314.

O'Quigley, J., Xu, R. Encyclopedia of Biostatistics (Vol. 2), chapter "Goodness-of-fit in survival analysis". Armitage, P., Colton, T., editors. Wiley; New York: 1998. p. 1731-1745.

O'Quigley, J., Xu, R. Handbook of Statistics in Clinical Oncology. In: Crowley, Hoering, editors. Explained variation and explained randomness for proportional hazards models. 3. Taylor & Francis Group, LLC; 2012. p. 487-503.

O'Quigley J, Xu R, Stare J. Explained randomness in proportional hazards models. Statistics in Medicine. 2005; 24:479–489. [PubMed: 15532086]

Preseley A, Tilahun A, Alonso A, Molenberghs G. An information-theoretic approach to surrogate-marker evaluation with failure time endpoints. Lifetime Data Analysis. 2011; 17:195–214. [PubMed: 20878357]

Russell M, Martier SS, Sokol R, Mudar P, Jacobson S, Jacobson J. Detecting risk-drinking during pregnancy: a comparison of four screening questionnaires. American Journal of Public Health. 1996; 84:1435–1439.

Schemper M, Stare J. Explained variation in survival analysis. Statistics in Medicine. 1996; 15:1999–2012. [PubMed: 8896135]

Xu, R. PhD Thesis. University of California; San Diego: 1996. Inference for the Proportional Hazards Model.

Xu R. Measuring explained variation in linear mixed effects models. Statistics in Medicine. 2003; 22:3527–3541. [PubMed: 14601017]

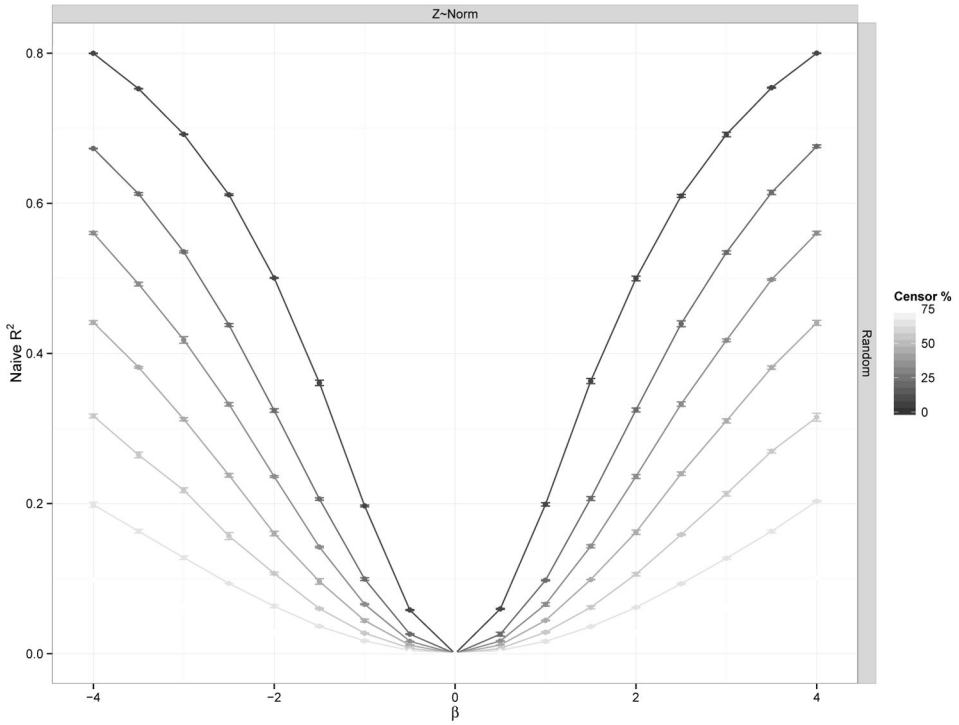

Figure 1.
Naive R^2 values using OLS with increasing amount of random censoring.

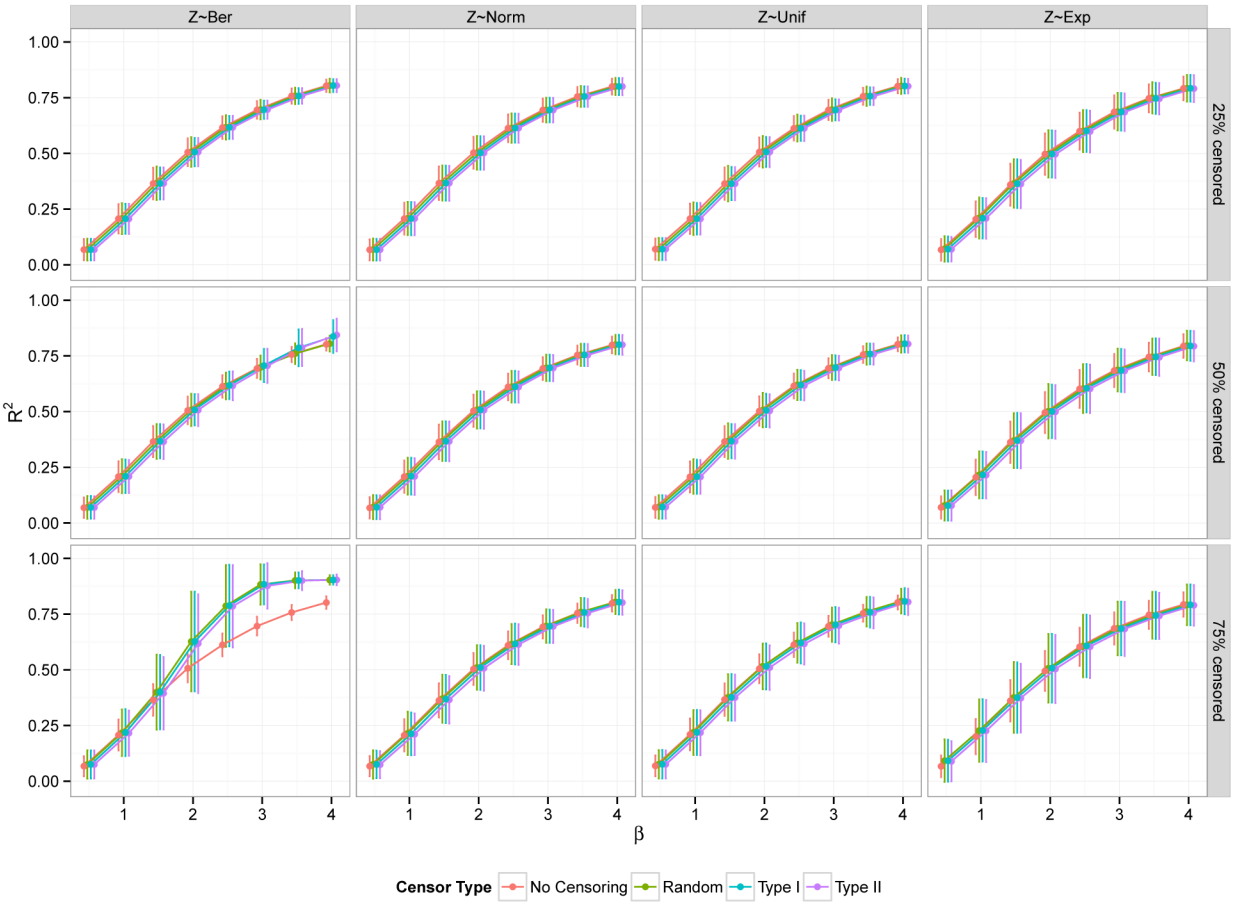

Figure 2.
Effects of covariate distributions and censoring on the R^2 measure.

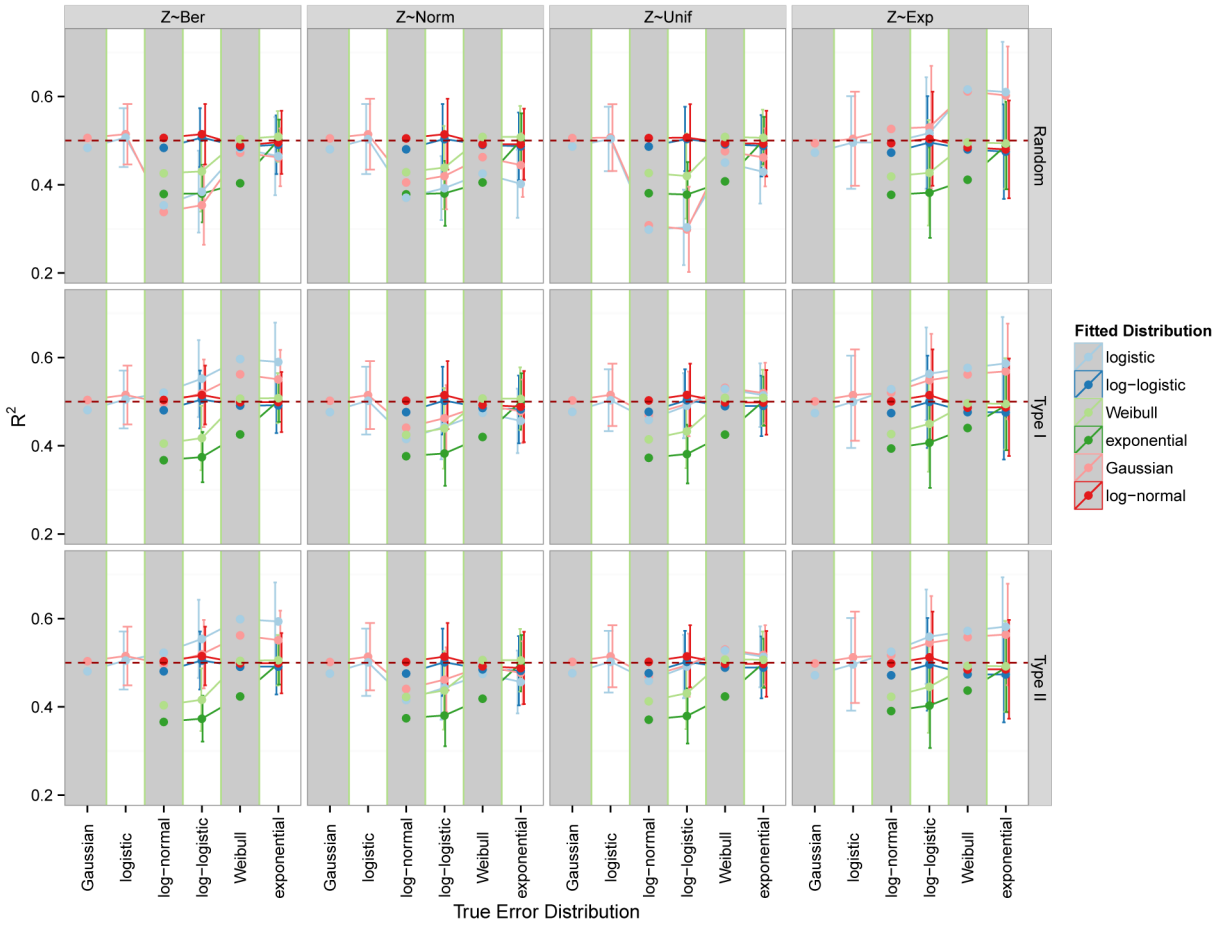

Figure 3.

R^2 performance under misspecified error distributions; $\Omega^2 = 0.5$ by construction, 25% censoring.

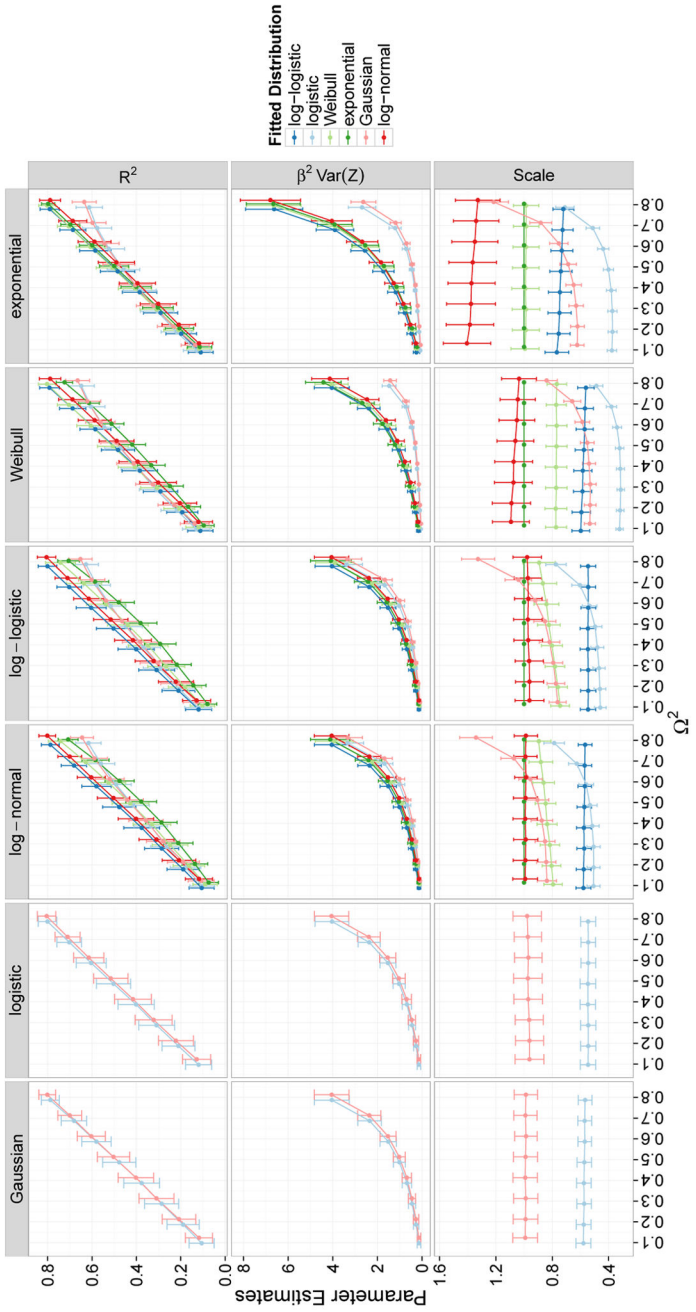

Figure 4.

R^2 performance under misspecified error distributions, for various values of Ω^2.

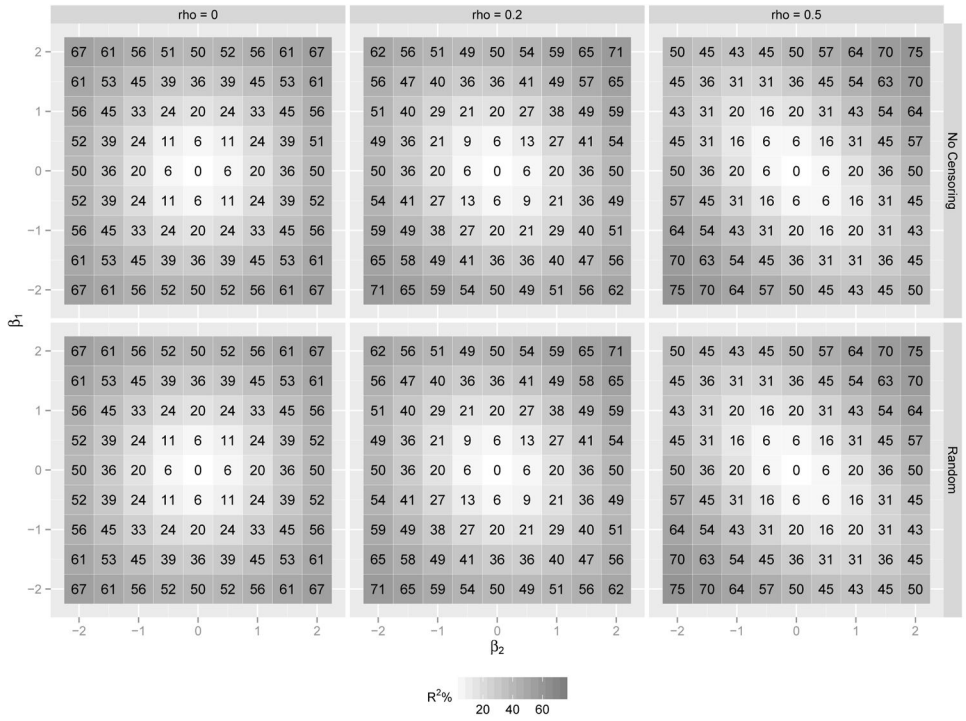

Figure 5.
R^2 values with bivariate normal covariates, 0% and 25% random censoring.

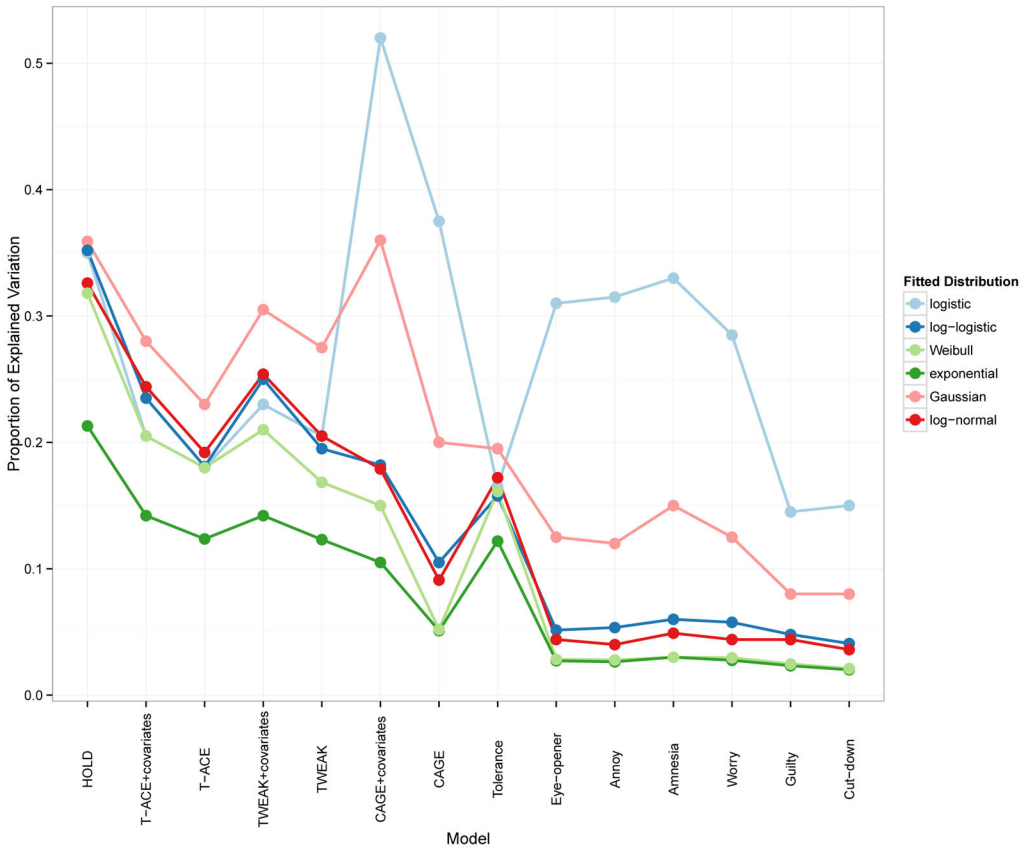

Figure 6.

R^2 values when different error distributions are fitted to the CIFASD data at conception.

Table 1

R^2 (standard deviation) under different true error distributions; $\Omega^2 = 0.5$ by construction, 25% censoring.

Error	Censoring	Covariate distribution							
		Bernoulli		Normal		Uniform		Exponential	
logistic	Random	0.507	(0.067)	0.504	(0.079)	0.504	(0.073)	0.496	(0.105)
	Type I	0.505	(0.066)	0.502	(0.077)	0.503	(0.070)	0.500	(0.105)
	Type II	0.505	(0.066)	0.501	(0.076)	0.502	(0.070)	0.496	(0.105)
log-logistic	Random	0.507	(0.067)	0.504	(0.079)	0.504	(0.073)	0.496	(0.105)
	Type I	0.505	(0.066)	0.502	(0.077)	0.503	(0.070)	0.500	(0.105)
	Type II	0.505	(0.066)	0.501	(0.076)	0.502	(0.070)	0.496	(0.105)
Weibull	Random	0.504	(0.060)	0.509	(0.074)	0.508	(0.063)	0.496	(0.101)
	Type I	0.507	(0.056)	0.507	(0.070)	0.509	(0.064)	0.497	(0.106)
	Type II	0.504	(0.057)	0.506	(0.070)	0.507	(0.064)	0.493	(0.106)
exponential	Random	0.502	(0.046)	0.501	(0.061)	0.500	(0.054)	0.489	(0.099)
	Type I	0.502	(0.048)	0.501	(0.064)	0.501	(0.056)	0.491	(0.101)
	Type II	0.498	(0.048)	0.499	(0.064)	0.499	(0.055)	0.488	(0.100)
Gaussian	Random	0.506	(0.066)	0.505	(0.074)	0.506	(0.068)	0.494	(0.101)
	Type I	0.504	(0.063)	0.502	(0.075)	0.503	(0.066)	0.501	(0.106)
	Type II	0.504	(0.063)	0.502	(0.075)	0.502	(0.066)	0.498	(0.105)
log-normal	Random	0.506	(0.066)	0.505	(0.074)	0.506	(0.068)	0.494	(0.101)
	Type I	0.504	(0.063)	0.502	(0.075)	0.503	(0.066)	0.501	(0.106)
	Type II	0.504	(0.063)	0.502	(0.075)	0.502	(0.066)	0.498	(0.105)

Table 2

R^2 values for various models of alcohol consumption around conception or in the most recent month during pregnancy in the Ukraine study, fitted with the log-normal error distribution.

Model	Conception	Most Recent
HOLD	0.326	0.413
T-ACE+covariates	0.244	0.307
T-ACE	0.192	0.279
TWEAK+covariates	0.254	0.316
TWEAK	0.205	0.293
CAGE+covariates	0.179	0.202
CAGE	0.091	0.140
Tolerance	0.172	0.251
Eye-opener	0.044	0.059
Annoy	0.040	0.045
Amnesia	0.049	0.064
Worry	0.044	0.049
Guilty	0.044	0.065
Cut-Down	0.036	0.045