Simulation Study: Cause-Specific Hazards

August 15, 2016

1 Motivation

We let T denote event time, $\varepsilon \in \{1, 2\}$ the corresponding cause and U the frailty, having density f_U, where U is independent of a randomized treatment X. We assume that the conditional cause-specific hazards of cause 1 and cause 2, respectively, are given as,

$$\begin{aligned}
\lambda_1(t \,|\, U, X) &= \lambda_1(t)U, \\
\lambda_2(t \,|\, U, X) &= \lambda_2(t)U \exp(\beta X), \quad \beta \neq 0.
\end{aligned}$$

Notice that $\lambda_1(t \,|\, U, X)$ is independent of X.

Now, we obtain the cumulative hazard function for cause 1 and cause 2, conditional on the frailty

$$\Lambda(t \,|\, U, X) = \big(\Lambda_1(t) + \Lambda_2(t) \exp(\beta X)\big)U,$$

resulting in the conditional survival,

$$S(t \,|\, U, X) = \exp\big(-\Lambda(t \,|\, U)\big) = \exp\Big(-\big(\Lambda_1(t) + \Lambda_2(t) \exp(\beta X)\big)U\Big).$$

Integrating out over the frailty distribution, we obtain the survival only conditional on X,

$$\begin{aligned}
S(t \,|\, X) = P(T > t \,|\, X) &= \int S(t \,|\, u, X) f_U(u) \,\mathrm{d}u \\
&= \int \exp\Big(-\big(\Lambda_1(t) + \Lambda_2(t) \exp(\beta X)\big)u\Big) f_U(u) \,\mathrm{d}u.
\end{aligned}$$

Specifically, we have

$$\begin{aligned}
P(T \leq t, \varepsilon = 1 \,|\, X) &= \int P(T \leq t, \varepsilon = 1 \,|\, u, X) \,\mathrm{d}u \\
&= \int \Big(\int_0^t S(t \,|\, u, X)\lambda_1(t \,|\, u, X) \,\mathrm{d}s\Big) f_U(u) \,\mathrm{d}u \\
&= \int_0^t \int S(t \,|\, u, X)\lambda_1(t) u f_U(u) \,\mathrm{d}u \,\mathrm{d}s \\
&= \int_0^t \int \exp\Big(-\big(\Lambda_1(t) + \Lambda_2(t) \exp(\beta X)\big)u\Big)\lambda_1(t) u f_U(u) \,\mathrm{d}u \,\mathrm{d}s.
\end{aligned}$$

Thus, the conditional cause-specific hazard of cause 1 given X becomes,

$$\begin{aligned}
\lambda_1(t \,|\, X) &= \frac{\frac{\partial}{\partial t} P(T \leq t, \varepsilon = 1 \,|\, X)}{P(T > t \,|\, X)} \\
&= \lambda_1(t) \frac{\int \exp\Big(-\big(\Lambda_1(t) + \Lambda_2(t) \exp(\beta X)\big)u\Big) u f_U(u) \,\mathrm{d}u}{\int \exp\Big(-\big(\Lambda_1(t) + \Lambda_2(t) \exp(\beta X)\big)u\Big) f_U(u) \,\mathrm{d}u} \qquad (1) \\
&= \lambda_1(t)\, \phi_U\Big(\exp\Big(-\big(\Lambda_1(t) + \Lambda_2(t) \exp(\beta X)\big)\Big)\Big),
\end{aligned}$$

where $\phi_U(v) = \frac{\partial}{\partial v} \log \mathbb{E}\big[\exp(vU)\big]$ is the derivative of the cumulant-generating function of U, i.e.

$$\phi_U(v) = \frac{\partial}{\partial v} \log \mathbb{E}\big[\exp(vU)\big] = \frac{\frac{\partial}{\partial v}\mathbb{E}\big[\exp(vU)\big]}{\mathbb{E}\big[\exp(vU)\big]} = \frac{\int \exp\big(vu\big) u f_U(u)\,\mathrm{d}u}{\int \exp\big(vu\big) f_U(u)\,\mathrm{d}u},$$

giving the third equality in (1) above. Note that (1) shows that the cause-specific hazard of cause 1 may be a function of X when frailty is not taken into account, even though it is independent of X when conditioning on the frailty.

1.1 Example: Gamma Frailty

We assume that the frailty follows a Gamma distribution, $U \sim \Gamma(\alpha, \beta)$. Note that the Laplace transform is then given as,

$$L_U(v) = \mathbb{E}\big[e^{-vU}\big] = \tfrac{\beta^\alpha}{(\beta+v)^\alpha},$$

so that the cumulant-generating function and its derivative becomes,

$$\log\, L_U(-v) = \alpha \log \beta - \alpha \log(\beta - v) \quad \Rightarrow \quad \phi_U(v) = \frac{\partial}{\partial v} \log\, L_U(-v) = \tfrac{\alpha}{\beta - v}.$$

Consequently, we obtain from (1) above, that

$$\lambda_1(t \,|\, X = x) = \tfrac{\lambda_1(t)\,\alpha}{\beta - e^{-\Lambda_1(t) - \Lambda_2(t)\exp(\beta x)}}.$$

2 Simulation Study

2.1 Simulation Competing Risks Data

We wish to simulate time T to any event ($\varepsilon \in \{1,2\}$) according to the distribution with cumulative distribution function

$$\begin{aligned} F(t) &\equiv P(T \leq t \,|\, U, X) = 1 - S(t \,|\, U, X) = 1 - \exp\big(-\Lambda(t \,|\, U)\big) \\ &= \exp\Big(-\big(\Lambda_1(t) + \Lambda_2(t)\exp(\beta X)\big)U\Big). \end{aligned}$$

This is done by use of $T \overset{\mathcal{D}}{=} F^{-1}(Z)$, where $Z \sim \text{unif}([0,1])$.

In addition, we simulate a censoring time C according to a Weibull distribution. The event $\varepsilon \in \{1,2\}$ is thus obtained whenever $T \leq C$.

Given a set of events, we determine which cause we are seeing from the cause-specific hazards. When we observe the event $\varepsilon \in \{1,2\}$ at time t under the values of $X = x$ and $U = u$, it is a case of cause 1 ($\varepsilon = 1$) with probability,

$$p_1(t, u, x) = \tfrac{\lambda_1(t \,|\, u, x)}{\lambda_1(t \,|\, u,x) + \lambda_2(t \,|\, u,x)},$$

and of cause 2 ($\varepsilon = 2$) with probability,

$$p_2(t, u, x) = 1 - p_1(t \,|\, u, x) = 1 - \tfrac{\lambda_1(t \,|\, u,x)}{\lambda_1(t \,|\, u,x) + \lambda_2(t \,|\, u,x)} = \tfrac{\lambda_2(t \,|\, u,x)}{\lambda_1(t \,|\, u,x) + \lambda_2(t \,|\, u,x)}.$$

That is, we simulate according to the Bernoulli distribution,

$$P\big(\varepsilon_t = 1 \,|\, \varepsilon_t \in \{1,2\}, u, x\big) = p_1(t,u,x), \quad P\big(\varepsilon_t = 2 \,|\, \varepsilon_t \in \{1,2\}, u, x\big) = p_2(t,u,x),$$

where ε_t denotes the event at time t. We let the cause-specific event times be $T^1 = T$ whenever $\varepsilon_T = 1$ and $T^2 = T$ whenever $\varepsilon_T = 2$.

2.2 Our Simulation Study

We simulate from a population of n individuals. We let

cause		cause-specific hazard
$\varepsilon = 1$	relapse	$\lambda_1(t \mid U, X) = \lambda_1(t)U$
$\varepsilon = 2$	death	$\lambda_2(t \mid U, X) = \lambda_2(t)U \exp(\beta X)$

We let $X_1, \ldots, X_n$ be outcomes from a binary variable ($p = 0.1627$), and we use $\beta = 0.7816$. Furthermore, the censoring times $C_1, \ldots, C_n$ are obtained from the Weibull distribution with shape parameter $\gamma = 1.3841$ and scale parameter $\lambda = 0.0271$ (note that our parametrization differs from the `R`-parametrization, where the scale is λ^{-1}). The cause 1 time-to-event and the cause 2 time-to-event are also assumed to follow a Weibull distribution, yielding the hazard function,

$$h(t) = \lambda\gamma(\gamma t)^{\gamma-1}. \tag{2}$$

Event times $T_1, \ldots, T_n$ are simulated as explained in Section 2.1. Event is then obtained as $1_{(T_i \leq C_i)}$, $i = 1, \ldots, n$, and mapped to either $\varepsilon = 1$ or $\varepsilon = 2$ as explained also in Section 2.1.

2.2.1 Obtaining Values of Parameter from Cancer Data

The parameters specified above are chosen such as to imitate a data set on leukaemia patients ($n = 1,715$).

We fit a Weibull distribution to the censoring times, the cause 1 times (relapse) and the cause 2 times (death), respectively (for the last two to be able to specify the hazard functions). From the hazard function in (2) that the cumulative hazard for the Weibull distribution becomes,

$$H(t) = \int_0^t h(s)\,\mathrm{d}s = \int_0^t \lambda\gamma(\gamma t)^{\gamma-1}\,\mathrm{d}s = (\lambda t)^\gamma,$$

so that,

$$\log H(t) = \gamma \log t + \gamma \log \lambda.$$

Thus, we can obtain estimates of γ and λ for each of the Weibull distributions by fitting a straight line through the log Nelson-Aalen estimates.

2.2.2 Tests

Ultimately, we aim to test if the general methods can detect the potential lack of model fit when data is simulated according to the scheme above, and frailty is not taken into account. Recall that the cause 2 hazard was a function of X, whereas the cause 1 hazard was not. We are primarily interested in the estimate $\hat{\beta}$ in the cause 1 cox regression over X (ignoring frailty). The tables show empirical averages over the estimation of the parameters using `coxph`, together with the fraction of p-values (H: $\beta = 0$) being significant. The boxplots show the distributions of the p-values of the model fits obtained from `cox.aalen`.

	1: $\hat{\beta}_U$ (sd)	**1:** $\hat{\beta}$ (sd)	**1:** $\mathbb{E}[1_{(p\leq 0.05)}]$	**2:** $\hat{\beta}_U$ (sd)	**2:** $\hat{\beta}$ (sd)	**2:** $\mathbb{E}[1_{(p\leq 0.05)}]$
true values	1.000	0.000	0.000	1.000	0.782	1.000
fit: X		-0.259 (0.254)	0.136		0.567 (0.120)	0.998
fit: $X + \log U$	1.007 (0.104)	-0.018 (-0.018)	0.040	1.000 (0.058)	0.794 (0.111)	1.000
sample size: 1000, no of simulations: 500						

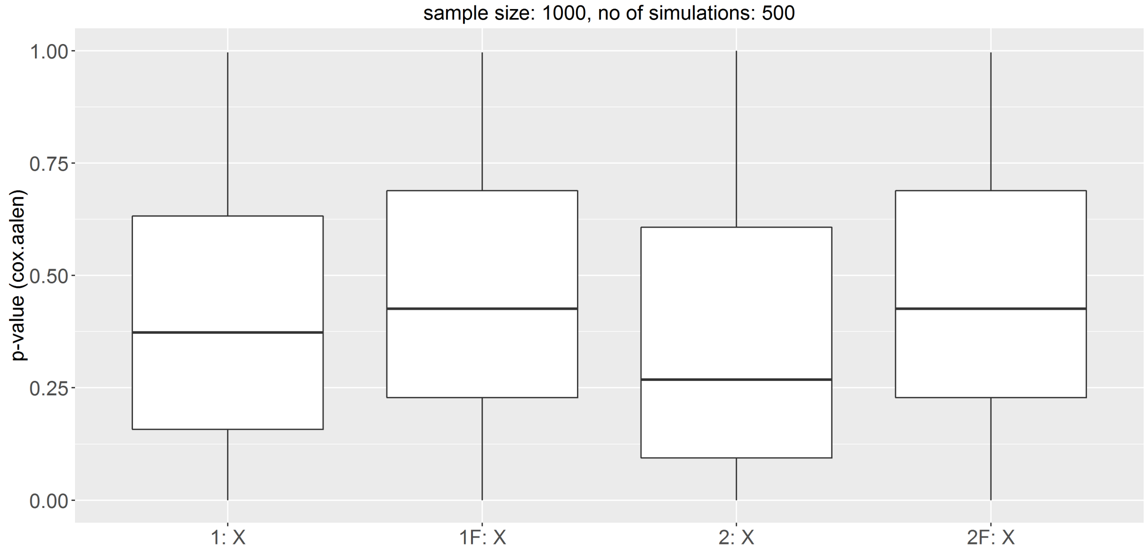

Figure 1: "1:X" refers to the cause 1 cox regression model over X, "1F:X" refers to the cause 1 cox regression model over $X + \log U$, "2:X" refers to the cause 2 cox regression model over X and "2F:X" refers to the cause 2 cox regression model over $X + \log U$.

	1: $\hat{\beta}_U$ (sd)	**1:** $\hat{\beta}$ (sd)	**1:** $\mathbb{E}[1_{(p\leq 0.05)}]$	**2:** $\hat{\beta}_U$ (sd)	**2:** $\hat{\beta}$ (sd)	**2:** $\mathbb{E}[1_{(p\leq 0.05)}]$
true values	1.000	0.000	0.000	1.000	0.782	1.000
fit: X		-0.243 (0.175)	0.246		0.561 (0.083)	1.000
fit: $X + \log U$	0.998 (0.067)	-0.011 (-0.011)	0.036	1.003 (0.041)	0.781 (0.080)	1.000
sample size: 2000, no of simulations: 500						

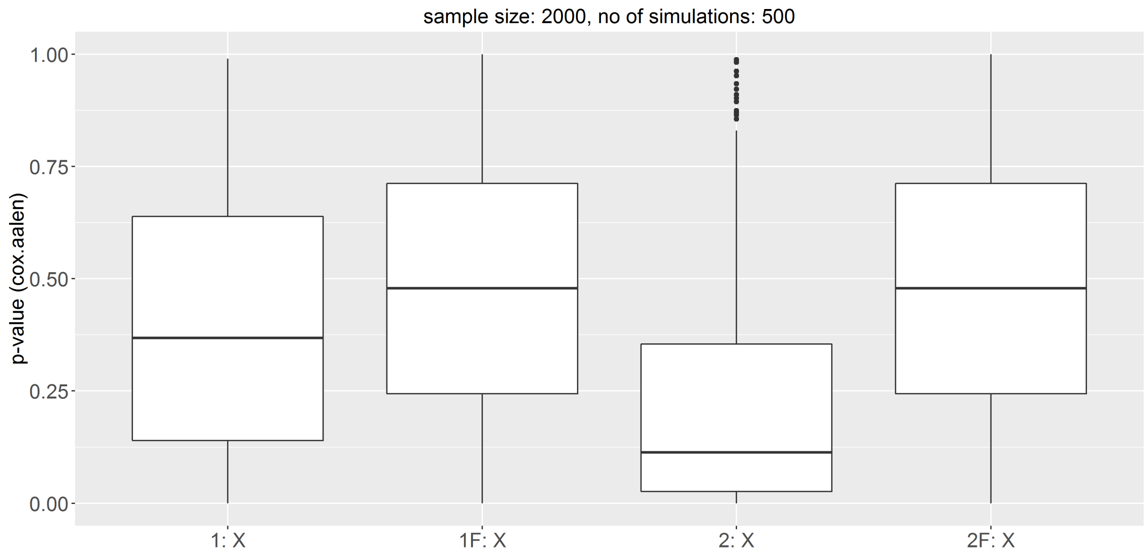

Figure 2: "1:X" refers to the cause 1 cox regression model over X, "1F:X" refers to the cause 1 cox regression model over $X + \log U$, "2:X" refers to the cause 2 cox regression model over X and "2F:X" refers to the cause 2 cox regression model over $X + \log U$.

	1: $\hat{\beta}_U$ (sd)	**1:** $\hat{\beta}$ (sd)	**1:** $\mathbb{E}[1_{(p\leq 0.05)}]$	**2:** $\hat{\beta}_U$ (sd)	**2:** $\hat{\beta}$ (sd)	**2:** $\mathbb{E}[1_{(p\leq 0.05)}]$
true values	1.000	0.000	0.000	1.000	0.782	1.000
fit: X		-0.240 (0.113)	0.574		0.559 (0.052)	1.000
fit: $X + \log U$	1.001 (0.042)	-0.005 (-0.005)	0.050	1.000 (0.027)	0.782 (0.050)	1.000
sample size: 5000, no of simulations: 500						

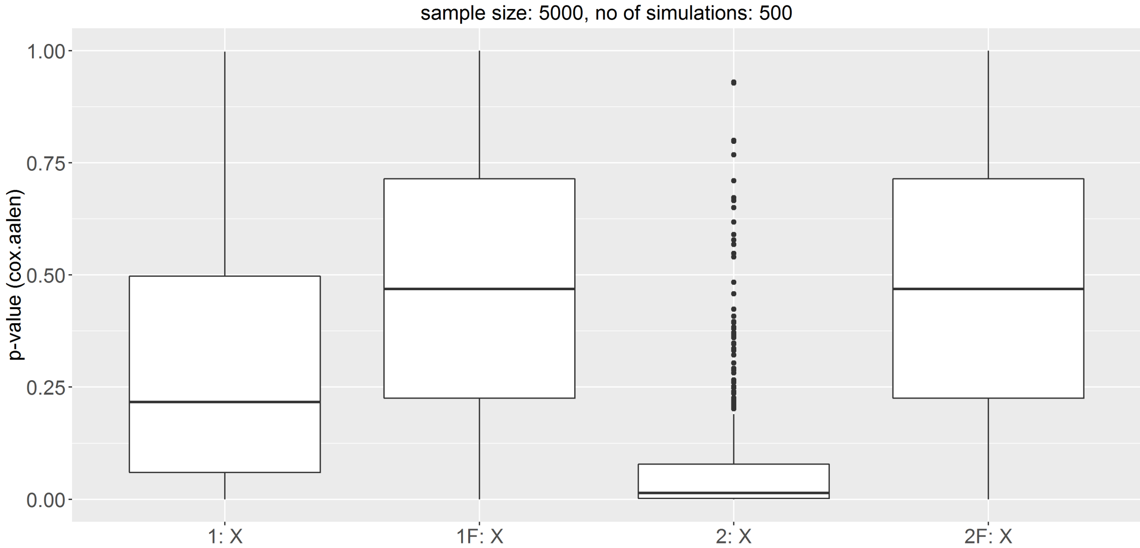

Figure 3: "1:X" refers to the cause 1 cox regression model over X, "1F:X" refers to the cause 1 cox regression model over $X + \log U$, "2:X" refers to the cause 2 cox regression model over X and "2F:X" refers to the cause 2 cox regression model over $X + \log U$.

	1: $\hat{\beta}_U$ (sd)	**1:** $\hat{\beta}$ (sd)	**1:** $\mathbb{E}[1_{(p\leq 0.05)}]$	**2:** $\hat{\beta}_U$ (sd)	**2:** $\hat{\beta}$ (sd)	**2:** $\mathbb{E}[1_{(p\leq 0.05)}]$
true values	1.000	0.000	0.000	1.000	0.782	1.000
fit: X		-0.241 (0.081)	0.876		0.556 (0.035)	1.000
fit: $X + \log U$	1.000 (0.029)	-0.004 (-0.004)	0.046	1.000 (0.018)	0.780 (0.035)	1.000
sample size: 10000, no of simulations: 500						

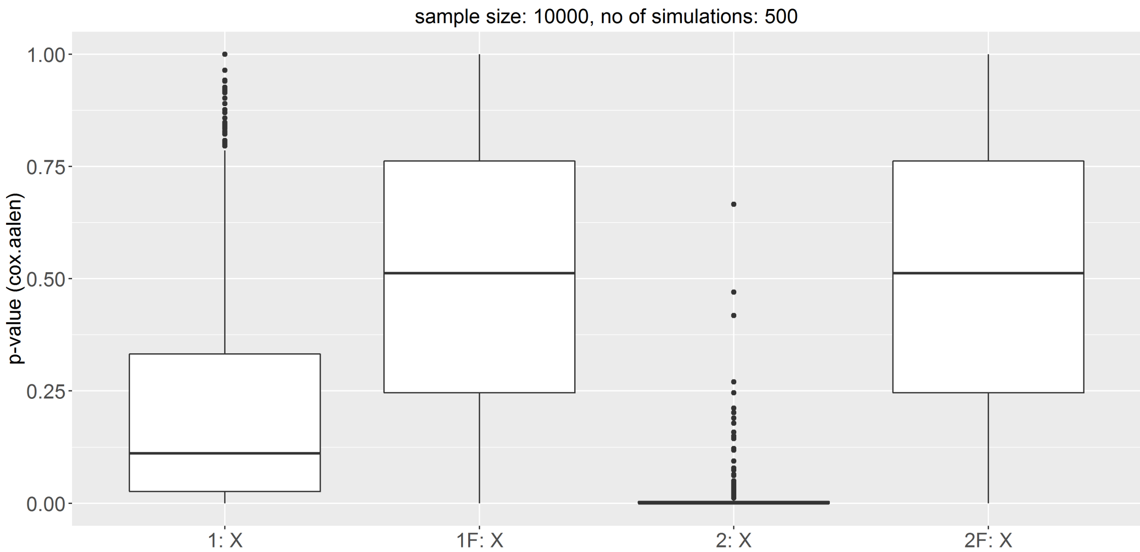

Figure 4: "1:X" refers to the cause 1 cox regression model over X, "1F:X" refers to the cause 1 cox regression model over $X + \log U$, "2:X" refers to the cause 2 cox regression model over X and "2F:X" refers to the cause 2 cox regression model over $X + \log U$.

	1: $\hat{\beta}_U$ (sd)	**1:** $\hat{\beta}$ (sd)	**1:** $\mathbb{E}[1_{(p\leq 0.05)}]$	**2:** $\hat{\beta}_U$ (sd)	**2:** $\hat{\beta}$ (sd)	**2:** $\mathbb{E}[1_{(p\leq 0.05)}]$
true values	1.000	0.000	0.000	1.000	0.782	1.000
fit: X		-0.240 (0.057)	0.994		0.558 (0.026)	1.000
fit: $X + \log U$	0.999 (0.021)	-0.005 (-0.005)	0.050	1.001 (0.013)	0.781 (0.024)	1.000
sample size: 20000, no of simulations: 500						

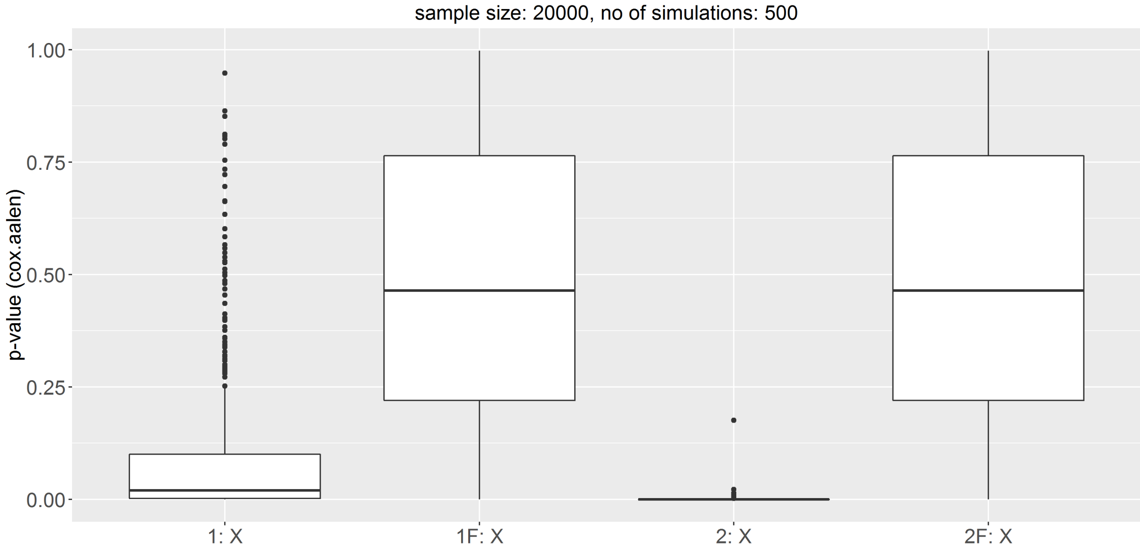

Figure 5: "1:X" refers to the cause 1 cox regression model over X, "1F:X" refers to the cause 1 cox regression model over $X + \log U$, "2:X" refers to the cause 2 cox regression model over X and "2F:X" refers to the cause 2 cox regression model over $X + \log U$.

	1: $\hat{\beta}_U$ (sd)	**1:** $\hat{\beta}$ (sd)	**1:** $\mathbb{E}[1_{(p\leq 0.05)}]$	**2:** $\hat{\beta}_U$ (sd)	**2:** $\hat{\beta}$ (sd)	**2:** $\mathbb{E}[1_{(p\leq 0.05)}]$
true values	1.000	0.000	0.000	1.000	0.782	1.000
fit: X		-0.235 (0.045)	1.000		0.559 (0.021)	1.000
fit: $X + \log U$	1.001 (0.017)	0.001 (0.001)	0.054	1.000 (0.010)	0.782 (0.021)	1.000
sample size: 30000, no of simulations: 500						

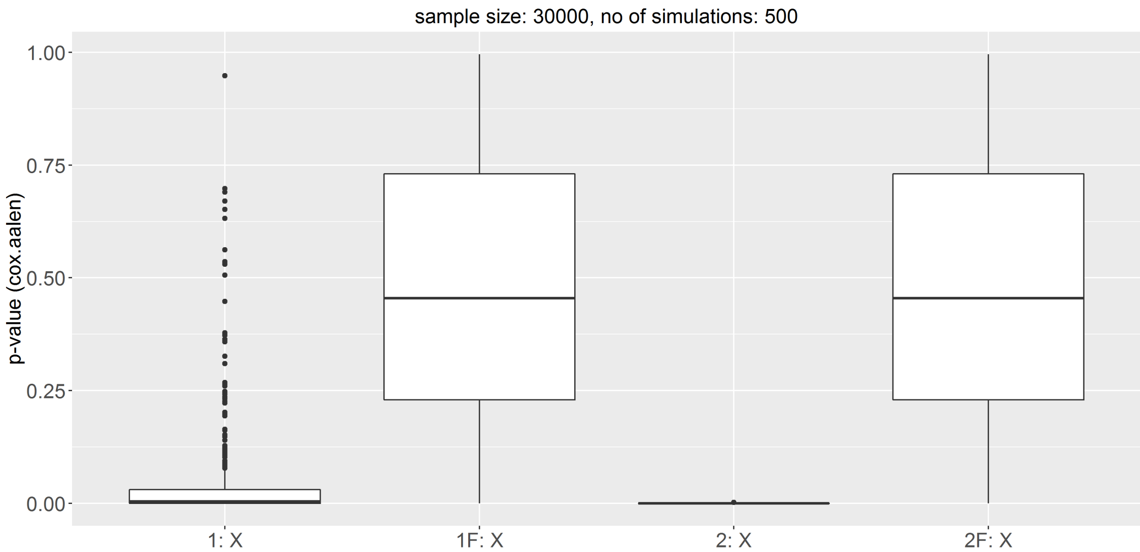

Figure 6: "1:X" refers to the cause 1 cox regression model over X, "1F:X" refers to the cause 1 cox regression model over $X + \log U$, "2:X" refers to the cause 2 cox regression model over X and "2F:X" refers to the cause 2 cox regression model over $X + \log U$.