Supplementary Material: 
Performance of two formal tests based on martingales residuals to check the proportional hazard assumption and the functional form of the prognostic factors in flexible parametric excess hazard models

CORALINE DANIELI$^{1-3,*}$, NADINE BOSSARD$^{1-2}$, LAURENT ROCHE$^{1-2}$, AURELIEN BELOT$^{4}$, ZOE UHRY$^{5,1-2}$, HADRIEN CHARVAT$^{6}$, LAURENT REMONTET$^{1-2}$, and the CENSUR Working Survival Group

$^1$Hospices Civils de Lyon, Service de Biostatistique-Bioinformatique, F-69003, Lyon, France; $^2$CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biotatistique-Santé, Université Lyon 1, F-69100, Villeurbanne, France; $^3$McGill University Health Center, Department of Epidemiology, Biostatistics and Occupational Health, H3A 1A2, Montreal, QC, Canada; $^4$Cancer Research UK Cancer Survival Group, Faculty of Epidemiology and Population Health, Department of NonCommunicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, U.K; $^5$Department of Chronic Diseases and Injuries, The French Public Health Agency, Saint Maurice, France; $^6$Epidemiology and Prevention Group, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan;

coraline.danieli@rimuhc.ca

*. To whom correspondence should be addressed.
1. Procedure used to generate $T_E$, the time to death from cancer

1.1 Simulation under the alternative hypothesis.

To assess the power of the test, $T_E$ must be generated from a model where age has a non-proportional effect. As recommended by Crowther and Lambert (2013), we simulated realistic data through a full parametric model which can be used in the general algorithm of simulation described by Crowther and Lambert (2013). $T_E$ was thus simulated from the following model

$$
\log [\lambda_E(t \mid age, \theta)] = \lambda_0(t, \chi) + \beta(t) \times age \quad (\text{Model 1})
$$

where the baseline excess mortality rate $\lambda_0(t, \chi)$ is a fractional polynomial:

$$
\lambda_0(t, \chi) = FP(t) = \beta_0 + \frac{\beta_1}{t+1} + \beta_2 \log (t + 1) + \beta_3 t + \beta_4 t^2
$$

and $\beta(t)$ is a cubic regression spline with 1 knot at one year:

$$
\beta(t) = \beta_5 + \beta_6 t + \beta_7 t^2 + \beta_8 t^3 + \beta_9 (t - 1)^3 I(t > 1)
$$

To simulate plausible effects with the three desired levels of non-proportionality, the values of $\beta_0$ to $\beta_9$ were chosen as follows (these choices are summarized in Table 1 in this document - http://www.biostatistics.oxfordjournals.org): i) we selected the 20 most frequent cancer sites from the common database of French cancer registries, ii) we fitted an excess hazard model on samples of each of these 20 real datasets (up to 10 years of follow-up), the excess mortality hazard $\lambda_E$ being specified as in Model 1, iii) for each cancer site, we measured the strength of the NPH effect with a particular measure defined as the area between $\beta(t)$ and the mean of $\beta(t)$ over time, and iv) we defined arbitrarily the “low NPH effect” as the $60^{th}$ percentile of this measure (which corresponds to stomach cancer), the “medium NPH effect” as the $80^{th}$ percentile (colon cancer), and the “high NPH effect” as the $90^{th}$ percentile (thyroid cancer). We chose percentiles over the $50^{th}$ so that the NPH effect could exist (i.e., simulation under $H_1$). Figure 1 (in this document) shows the shape of the three NPH effects $\beta(t)$ according to the time since diagnosis. Furthermore, the power of the test was deemed dependent on the number of deaths observed.
Martingale residuals for checking excess hazard model assumptions

in the simulated cohort, this number being dependent on the baseline hazard level and on the NPH level. So, to analyze the specific effect of the NPH levels, we controlled the baseline excess mortality hazard and fixed its shape and strength using parameters $\beta_0$ to $\beta_4$ observed in colon cancer.

1.2 Simulation under the null hypothesis.

Finally, to assess the size, $T_E$ was generated from the following proportional hazard model:

$$\log \left[ \lambda_E (t|\text{age}, \theta) \right] = \beta_0 + \frac{\beta_1}{t+1} + \beta_2 \log (t + 1) + \beta_3 t + \beta_4 t^2 + \beta_5 \times \text{age} \quad \text{(Model 2)}$$

where the values of $\beta_0$ to $\beta_5$ were estimated from colon cancer data (Table 1 in this document).

After determining the parameters of the model, $T_E$ was generated using the inverse transform approach described by Crowther and Lambert (2013) for time-dependent effects: the parametric form of the model allows obtaining the cumulative distribution which is then numerically inverted.

2. Simulation study to assess the performance of the test of the FF of a covariate

This simulation study is largely based on the elements already presented in the previous section. Thus, we present here only the elements that differ from this previous section. We assessed the size and the power of the new test designed to check the FF of a covariate (called thereafter “the new FF test”). This assessment was performed in the particular case where hypothesis $H_0$ is “the functional form of the covariate is linear”. The performance of this test was estimated according to the sample size and the strength of the non-linearity (NLIN) effect (low, medium, high) under the alternative hypothesis $H_1$. To obtain plausible effects with the three levels of NLIN strength, we operated as follows (Table 1 in this document): i) for each of the 20 most frequent cancer sites, we fit the following model

$$\log \left[ \lambda_E (t|\text{age}, \theta) \right] = \lambda_0 (t, \chi) + g (\text{age})$$

being a fractional polynomial and $g (\text{age})$ a cubic spline with one knot at the mean age, $g (\text{age}) =$
\[ \beta_5 \times age + \beta_6 \times age^2 + \beta_7 \times age^3 + \beta_8 \times (age - mean)^3 I (age > mean) \]
i) we measured the strength of the NLIN effect with a specific measure similar to that of the new PH test but adapted for linearity; this corresponds to the area between \( g(age) \) and the mean of \( g(age) \) over the age range; iii) from the percentiles of this measure, we defined arbitrarily a “low”, a “medium”, and a “high” NLIN effect that correspond, respectively, to colon, cervix, and prostate cancers.
Figure 1. Shape of the three simulated NPH effects $\beta(t)$: circles for low NPH effect, triangles for medium NPH effect, and crosses for high NPH effect.
### Table 1. Summary of the simulation design

| Test  | Criterion | Simulated Model used for Source of Source of NPH/NLIN Strength effect |
|-------|-----------|-------------------|----------------|------------------|------------------|
|       | Data under | $\lambda_E(t, \text{age})$ | $\beta_0-\beta_4$ values | $\beta_5-\beta_9$ values | $\log [\lambda_E(t, \text{age})]$ in analysis |
| PH    | Power $H_1$ | $FP(t) + \beta(t) \times \text{age}$ | Colon cancer | Stomach cancer | Low | $\sim QS(t) + \alpha_5 \times \text{age}$ |
| PH    | Power $H_1$ | $FP(t) + \beta(t) \times \text{age}$ | Colon cancer | Colon cancer | Medium | $\sim QS(t) + \alpha_5 \times \text{age}$ |
| PH    | Power $H_1$ | $FP(t) + \beta(t) \times \text{age}$ | Colon cancer | Thyroid cancer | High | $\sim QS(t) + \alpha_5 \times \text{age}$ |
| PH    | Size $H_0$ | $FP(t) + \beta_5 \times \text{age}$ | Colon cancer | Colon cancer | None | $\sim QS(t) + \alpha_5 \times \text{age}$ |
| FF    | Power $H_1$ | $FP(t) + g(\text{age})$ | Colon cancer | Colon cancer | Low | $\sim QS(t) + \alpha_5 \times \text{age}$ |
| FF    | Power $H_1$ | $FP(t) + g(\text{age})$ | Colon cancer | Cervix cancer | Medium | $\sim QS(t) + \alpha_5 \times \text{age}$ |
| FF    | Power $H_1$ | $FP(t) + g(\text{age})$ | Colon cancer | Prostate cancer | High | $\sim QS(t) + \alpha_5 \times \text{age}$ |
| FF    | Size $H_0$ | $FP(t) + \beta_5 \times \text{age}$ | Colon cancer | Colon cancer | None | $\sim QS(t) + \alpha_5 \times \text{age}$ |

(1) Fractional Polynomial: $FP(t) = \beta_0 + \frac{\beta_1}{t+1} + \beta_2 \log (t+1) + \beta_3 t + \beta_4 t^2$

$\beta(t) = (\beta_5 + \beta_6 t + \beta_7 t^2 + \beta_8 t^3 + \beta_9 (t-1)^3) I(t > 1)$

$g(\text{age}) = \beta_5 \times \text{age} + \beta_6 \times \text{age}^2 + \beta_7 \times \text{age}^3 + \beta_8 \times (\text{age} - \text{mean})^3 I(\text{age} > \text{mean})$

(2) Quadratic regression spline: $QS(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 (t-1)^2 I(t > 1) + \alpha_4 (t-5)^2 I(t > 5)$
3. Approximation of the limiting distribution of the score process under the null hypothesis.

Most of this section comes from Lin and Spiekerman (1996). Here, we present the main steps, starting from the score process, to derive the process that was simulated to approximate the limiting distribution of the score process.

Let $\theta_0$ be the true values of the parameters. A Taylor expansion of $U(\theta, t)$ (the score process evaluated at time $t$) in the neighborhood of $\theta_0$ gives:

$$U(\theta, t) = U(\theta_0, t) - nI(\theta_0, t)(\theta - \theta_0) + o_p(1) \quad (3.1)$$

where $I(\theta, t)$ is a $p \times p$ matrix ($p$ equals to the number of parameters in the model) with the element $I_{kj}(\theta, t)$ (i.e., the element in the $k^{th}$ row and $j^{th}$ column) equals to:

$$-\frac{1}{n} \frac{\partial U_k}{\partial \theta_j}(\theta, t)$$

The matrix $I$ evaluated at $(\theta_0, \tau)$ is related to the Fisher information matrix and is invertible. Evaluating (3.1) at $(\hat{\theta}, \tau)$ and remarking that $U(\hat{\theta}, \tau)$ corresponds to the score of the likelihood function evaluated at the maximum likelihood estimate (and thus is equal to the zero vector), we obtain:

$$nI(\theta_0, \tau)(\hat{\theta} - \theta_0) = U(\theta_0, \tau) + o_p(1)$$

$$\left(\hat{\theta} - \theta_0\right) = \frac{1}{n} I^{-1}(\theta_0, \tau) U(\theta_0, \tau) + o_p(1) \quad (3.2)$$

Then, combining equation (3.1) and (3.2) and evaluating $U$ at $\hat{\theta}$ and time $t$,

$$U(\hat{\theta}, t) = U(\theta_0, t) - I(\theta_0, t) I^{-1}(\theta_0, \tau) U(\theta_0, \tau) + o_p(1) \quad (3.3)$$

In equation (3.3), the distribution of the components of the score process $U_k(\theta_0, t)$ is unknown. Indeed, the $k^{th}$ component of the score process at $\theta_0$ is:

$$U_k(\theta_0, t) = \sum_i Z_{ki} \int_0^t \frac{\lambda_E(u|Z_i, \theta_0)}{\lambda_E(u|Z_i, \theta_0) + \lambda_F(u_i + u, z_i)} dM_i(u)$$
with
\[ M_i(t) = N_i(t) - \int_0^t Y_i(u) \lambda_E(u|Z_i, \theta_0) du - \int_0^t Y_i(u) \lambda_P(a_i + u, \bar{z}_i) du \]

The distribution of \( M_i \) is unknown and, similarly to Lin and others (1993) and Lin and Spiekerman (1996), we replace \( M_i \) by a similar process \( M_i G_i \), with \( G_i \) following a normal distribution. This process shares characteristics of the process \( M_i \), including mean and variance (Fleming and Harrington, 1991; Lin and others, 1993).

Thus, \( U_k(\theta_0, .) \) will be approximated by the \( k^{th} \) component of \( D_1(\theta, .) \) defined as follows:
\[
\sum_i \left[ Z_{ki} \int_0^t \frac{\lambda_E(u|Z_i, \theta)}{\lambda_E(u|Z_i, \theta) + \lambda_P(a_i + u, \bar{z}_i)} dN_i(u) G_i \right]
\] (3.4)

Here, we focus on the score process. In this context, \( n^{-1/2} U \) is a special case of the process \( W_z^{(2)} \) (see the end of Section 2.2.2 of the article). Finally, combining equations (3.3) and (3.4), and approximating \( \theta_0 \) by \( \hat{\theta} \), the approximation of the distribution of \( \left\{ n^{-1/2} \hat{U} \right\} \) is performed by generating a ‘large’ number of realizations of the following process:
\[
n^{-1/2} \left( D_1(\hat{\theta}, t) - I(\hat{\theta}, t) I(\hat{\theta}, \tau)^{-1} D_1(\hat{\theta}, \tau) \right)
\] (3.5)

To check PH assumption of the \( k^{th} \) covariate, the “observed” test statistic value of the corresponding dataset is:
\[
\hat{T}_S = \sup \left| n^{-1/2} U_k(\hat{\theta}, t) \right|
\] (3.6)

The p-value is then approximated by the proportion of cases in which the absolute maximum of the simulated Gaussian Processes (supremum over time \( t \) of the absolute value of the \( k^{th} \) component of process (3.5), i.e., replacing \( U_k \) in (3.6) with the \( k^{th} \) component of (3.5)) is higher than \( \hat{T}_S \).

References

