Supplemental material for

“Proportionally Less Difficult?: Reevaluating Keele’s ‘Proportionally Difficult’”

Shawna K. Metzger
Michigan State University
metzge69@msu.edu

Appendix A
PH Test: Actual vs. Approximation

I. The Test Statistic

Grambsch and Therneau's actual test statistic for their proportional hazards (PH) test is equal to (Therneau and Grambsch 2000, 132):[1]

$$T(G) = \tilde{\theta}' Q \tilde{\theta} \tag{1}$$

When the PH assumption holds, $T(G)$ is distributed asymptotically χ^2. G is a generic identifier to represent the selected time transform, $g(t)$, and (Therneau and Grambsch 2000, 132)

$$\tilde{\theta} = Q^{-1} \sum_k G_k s_k \tag{2}$$

$$Q = \sum_k G_k \hat{V}_k G_k' - \left(\sum_k G_k \hat{V}_k\right)\left(\sum_k \hat{V}_k\right)^{-1}\left(\sum_k G_k \hat{V}_k\right)' \tag{3}$$

k is the kth event time ($0 < t_1 < \cdots < t_k < t_K$). All of the matrices are specific to the kth event time:

- G_k: $J \times J$ diagonal matrix with $g(t_k) - \overline{g(t)}$ on the diagonal
- s_k: $J \times 1$ vector containing the unscaled Schoenfeld residuals at k, for all $j = 1, 2, \ldots J$ covariates
- $\hat{V}_k$: the $J \times J$ variance-covariance matrix for the Cox model's estimates at k. $\hat{V}_k$'s formula appears in an earlier chapter (Therneau and Grambsch 2000, 40 [Eq. 3.7]).

II. Implementation: Actual Test

Therneau and Grambsch note that there are surprisingly familiar ways to obtain $\tilde{\theta}$ and $T(G)$. To quote the relevant paragraph in full (2000, 132):

> The estimator and test statistic can also be derived from standard partial likelihood arguments: $\tilde{\theta}$ is a one-step Newton-Raphson algorithm update starting from $(\beta, \theta) = (\hat{\beta}, 0)$ and $T(G)$ is the Rao efficient score test of $H_0 : \theta = 0$. The asymptotic distribution

[1] Where possible in this appendix, I use the same notation as Therneau and Grambsch (2000).

> of $T(G)$ under H_0 follows from the properties of score processes of partial likelihoods using martingale asymptotics.

`survival`'s new implementation of the actual PH test uses this logic: it modifies the C routines underlying `survival::coxph`, and the modified routine only iterates once (Therneau 2021, lines 16–17, 142–143).

A score test, also known as a Rao efficient score test or a Lagrange multiplier (LM) test, takes the form (Greene 2003, 489–90)

$$LM = U\mathcal{I}^{-1}U'$$

where U is the score vector, as a row, and $\mathcal{I}$ is the information matrix.

A covariate's entry in the score vector is equal to the sum of its Schoenfeld residuals, making U particularly easy to compute in a Cox model context (Therneau and Grambsch 2000, 40, 85). It also makes the equivalence between Eq. (1) and the score test more obvious.[2] Therneau's documentation for `cox.zph`'s updated routine explains how the test is implemented (2021, lines 2–35):[3]

> The simplest test of proportional hazards is to use a time dependent coefficient $\beta(t) = a + bt$. Then $\beta(t)x = ax + b(tx)$, and the extended coefficients a and b can be obtained from a Cox model with an extra 'fake' covariate tx. More generally, replace t with some function $g(t)$, which gives rise to an entire family of tests. An efficient assessment of this extended model can be done using a score test.
>
> - Augment the original variables $x_1, \ldots, x_J$ with J new ones $g(t)x_1, \ldots, g(t)x_J$
> - Compute the first and second derivatives U and $\mathcal{I}$ of the Cox model at the starting estimate of $(\hat{\beta}, 0)$; prior covariates at their prior values, and the new covariates at 0. No iteration is done. [...]

[2] See this appendix, Section III.A for full demonstration.

[3] I have shifted some notation in this excerpt so as to be consistent with my earlier usage.

- By design, the first J elements of U will be zero. Thus the first iteration of the new coefficients, and the score tests for them, are particularly easy.

The information or Hessian matrix for a Cox model is

$$\sum_{k \epsilon K} \hat{V}(t_k) = \sum_k \hat{V}_k$$

where $\hat{V}_k$ is the variance matrix of the weighted covariate values, over all subjects at risk at time t_k. Then the expanded information matrix for the score test is

$$\mathcal{I} = \begin{pmatrix} \mathcal{I}_1 & \mathcal{I}_2 \\ \mathcal{I}_2' & \mathcal{I}_3 \end{pmatrix}$$

$$\mathcal{I}_1 = \sum \hat{V}(t_k)$$

$$\mathcal{I}_2 = \sum \hat{V}(t_k) g(t_k)$$

$$\mathcal{I}_3 = \sum \hat{V}(t_k) g^2(t_k)$$

The actual score test does all these calculations out, in full. `survival` 3.0-10 and after implements them in C, for speed and precision reasons (`zph1.c`/`zph2.c`). The *J*-length expanded portion of the score vector will be equal to

$$U = \left(\sum_k s_{j=1}\, g(t_k), \cdots, \sum_k s_{j=J}\, g(t_k) \right)$$

where $s_{j=c}$ is the *K*-length vector of Schoenfelds for the *c*th covariate.

III. Implementation: Approximation

The approximation makes a key simplifying assumption about $\hat{V}(t_k)$. It stems from the observation that $\hat{V}(t_k)$ usually "changes slowly, and is quite stable until the last few death times" (Therneau and Grambsch 2000, 133–34). If we are willing to assume $\hat{V}(t_k)$ is exactly constant—i.e., the variance is constant across all *t*s—then $\mathcal{I}_2 = 0$ and $\mathcal{I}_3 = \sum \hat{V}(t_k) \sum g^2(t_k)$ (Therneau 2021, lines 38–41).

As an additional implication, because

$$\mathcal{I} = \sum \hat{V}(t_k)$$

it also suggests a reasonable substitution for $\hat{V}(t_k)$: the *average* of $\hat{V}(t_k)$ across all the failure times, $\bar{V} = d^{-1} \sum \hat{V}(t_k) = d^{-1}\mathcal{I}$. The substitution is desirable because $\hat{V}(t_k)$ "may be unstable, particularly near the end of follow up when the number of subjects in the risk set is not much larger than the number of rows for" $\hat{V}(t_k)$ (Therneau and Grambsch 2000, 133–34). By contrast, $\bar{V}$'s value is more stable.

With the assumption and the $\bar{V}$ substitution it implies in hand, we can now show how the approximation's formula for covariate *j* (Therneau and Grambsch 2000, 134 [Eq. 6.6])

$$T = \frac{\left\{\sum_k s_k^* \left[g(t_k) - \overline{g(t)}\right]\right\}^2}{d\hat{V}_{\hat{\beta}_j} \sum_k \left(\left[g(t_k) - \overline{g(t)}\right]^2\right)} \tag{4}$$

follows from the full, actual formula (Eqs. (1)–(3)), where s_k^* is the scaled Schoenfeld residual for the *k*th failure time.

To simplify matters, I use a Cox model with no strata and a single covariate ($j = J = 1$) to work through the formula.[4] A single covariate means the Schoenfeld residuals will be a $K \times 1$ column vector; I drop any covariate-related subscripts for the Schoenfelds to streamline, leaving only the failure time-related subscripts (*k*).

[4] The broad logic generalizes to $J > 1$. In this situation, U_j is subset to the first J entries, plus the entry corresponding to the x_j*$g(t)$ interaction, producing a vector of length $J + 1$. $\mathcal{I}_j$ is subset similarly to produce a $(J + 1) \times (J + 1)$ matrix: all of $\mathcal{I}_1$ stays, in addition to the row and column corresponding to x_j*$g(t)$ from the expanded portion of $\mathcal{I}$ (`survival` 3.2.11, `cox.zph.R`, lines 123–139).

A. Score Test Notation

For the score test, we will need (1) the score vector and (2) the information matrix's inverse. The expanded score vector's value is the same when $J = 1$, in both the actual calculation and the approximation:[5]

$$U_j = \left[0 \quad \sum_k s_k \, g(t_k)\right]$$

The information matrix is not the same across the two calculations. For the approximation, we start with

$$\mathcal{I}_j = \begin{bmatrix} \sum \hat{V}_j(t_k) & 0 \\ 0 & \sum \hat{V}_j(t_k) \sum g^2(t_k) \end{bmatrix}$$

Once we make the $\bar{V}$ substitution, we obtain

$$\mathcal{I}_j = \begin{bmatrix} \frac{\sum \hat{V}(t_k)}{d} & 0 \\ 0 & \frac{\sum \hat{V}(t_k)}{d} \sum g^2(t_k) \end{bmatrix}$$

$\sum \hat{V}(t_k) = \mathcal{I}_{jj} = 1/\hat{V}_{\hat{\beta}_j}$, where $\hat{V}_{\hat{\beta}_j}$ is $\hat{\beta}_j$'s estimated variance from the original Cox model. Making this last substitution gives us

$$\mathcal{I}_j = \begin{bmatrix} \frac{1}{d\hat{V}_{\hat{\beta}_j}} & 0 \\ 0 & \frac{1}{d\hat{V}_{\hat{\beta}_j}} \sum g^2(t_k) \end{bmatrix}$$

We can now take this matrix's inverse. Doing so and simplifying yields

$$\mathcal{I}_j^{-1} = \frac{d\hat{V}_{\hat{\beta}_j}}{\sum g^2(t_k)} \begin{bmatrix} \sum g^2(t_k) & 0 \\ 0 & 1 \end{bmatrix}$$

[5] Differences emerge when $J > 1$ for the covariate-specific calculations (vs. the global test statistic), as a byproduct of the approximation's simplifying assumption. The assumption impacts the link between the scaled and unscaled Schoenfelds (Therneau and Grambsch 2000, 131 [actual], 134 [approximation]).

The final step is the actual score test calculation:

$$LM_j = U_j \mathcal{I}_j^{-1} U_j'$$

$$LM_j = \frac{d\hat{V}_{\hat{\beta}_j}}{\sum g^2(t_k)} \left[0 \sum\nolimits_k s_k\, g(t_k)\right] \begin{bmatrix} \sum g^2(t_k) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \sum_k s_k\, g(t_k) \end{bmatrix}$$

$$LM_j = \frac{d\hat{V}_{\hat{\beta}_j}}{\sum g^2(t_k)} \left[0 \sum\nolimits_k s_k\, g(t_k)\right] \begin{bmatrix} 0 \\ \sum_k s_k\, g(t_k) \end{bmatrix}$$

$$LM_j = \frac{d\hat{V}_{\hat{\beta}_j} \{\sum_k s_k\, g(t_k)\}^2}{\sum g^2(t_k)} \tag{5}$$

Therneau and Grambsch demean $g(t_k)$ to center its value at zero because "[t]he inverse of the matrix will be more numerically stable…and this does not change the test statistic['s value]" (Therneau 2021, lines 34–35). Inserting demeaned $g(t_k)$ gives us:

$$LM_j = \frac{d\hat{V}_{\hat{\beta}_j} \left\{\sum_k s_k \left[g(t_k) - \overline{g(t)}\right]\right\}^2}{\sum_k \left(\left[g(t_k) - \overline{g(t)}\right]^2\right)} \tag{6}$$

This expression does not yet match Eq. (4) because the latter uses the scaled Schoenfeld residuals (s_k^*) instead of Eq. (6)'s unscaled Schoenfelds (s_k). We can express the scaled Schoenfelds in terms of the unscaled Schoenfelds if we utilize our $\bar{V}$ substitution. For covariate *j* when *J* = 1, it will be (Therneau and Grambsch 2000, 134):

$$s_j^* = d\hat{V}_{\hat{\beta}_j} s_j$$

$$s_j = \frac{s_j^*}{d\hat{V}_{\hat{\beta}_j}}$$

Using this information, we can make a final set of substitutions and simplifications:

$$LM_j = \frac{d\hat{V}_{\hat{\beta}_j} \left\{\sum_k \frac{s_k^*}{d\hat{V}_{\hat{\beta}_j}} \left[g(t_k) - \overline{g(t)}\right]\right\}^2}{\sum_k \left(\left[g(t_k) - \overline{g(t)}\right]^2\right)}$$

$$LM_j = \frac{d\hat{V}_{\hat{\beta}_j} \left\{\frac{1}{d\hat{V}_{\hat{\beta}_j}} \sum_k s_k^* \left[g(t_k) - \overline{g(t)}\right]\right\}^2}{\sum_k \left(\left[g(t_k) - \overline{g(t)}\right]^2\right)}$$

$$LM_j = \frac{\frac{d\hat{V}_{\hat{\beta}_j}}{d^2\left(\hat{V}_{\hat{\beta}_j}\right)^2}\left\{\sum_k s_k^* \left[g(t_k) - \overline{g(t)}\right]\right\}^2}{\sum_k \left(\left[g(t_k) - \overline{g(t)}\right]^2\right)}$$

$$LM_j = \frac{\frac{1}{d\hat{V}_{\hat{\beta}_j}}\left\{\sum_k s_k^* \left[g(t_k) - \overline{g(t)}\right]\right\}^2}{\sum_k \left(\left[g(t_k) - \overline{g(t)}\right]^2\right)}$$

$$\boxed{LM_j = \frac{\left\{\sum_k s_k^* \left[g(t_k) - \overline{g(t)}\right]\right\}^2}{d\hat{V}_{\hat{\beta}_j} \sum_k \left(\left[g(t_k) - \overline{g(t)}\right]^2\right)}}$$

The boxed expression and Eq. (4) now match.

B. $\tilde{\theta}$/Q Notation

We can also use Therneau and Grambsch's definitions for $\tilde{\theta}$ and Q to obtain Eq. (4), which may not be immediately obvious when everything is written in matrix form. Note that our simplifying assumption about $\hat{V}(t_k)$'s value makes Q (Eq. (3)) equal to (Therneau and Grambsch 2000, 134)

$$Q = \frac{\sum G_k \mathcal{I} G_k'}{d}$$

Continuing our previous example of $j = J = 1$ for simplicity, we obtain

$$Q_j = \frac{\mathcal{I}_{jj} \sum g^2(t_k)}{d}$$

$$Q_j = \frac{\sum g^2(t_k)}{d\hat{V}_{\hat{\beta}_j}}$$

And, from Eq. (2),

$$\tilde{\theta}_j = \frac{d \sum_k s_k \, g(t_k)}{\mathcal{I}_{jj} \sum g^2(t_k)}$$

$$\tilde{\theta}_j = \frac{d\hat{V}_{\hat{\beta}_j} \sum_k s_k \, g(t_k)}{\sum g^2(t_k)}$$

The final test statistic's expression is then

$$T_j(G) = \tilde{\theta}_j{}' Q_j \tilde{\theta}_j$$

$$T_j(G) = \frac{d\hat{V}_{\hat{\beta}_j} \sum_k s_k \, g(t_k)}{\sum g^2(t_k)} * \frac{\sum g^2(t_k)}{d\hat{V}_{\hat{\beta}_j}} * \frac{d\hat{V}_{\hat{\beta}_j} \sum_k s_k \, g(t_k)}{\sum g^2(t_k)}$$

$$\boxed{T_j(G) = \frac{d\hat{V}_{\hat{\beta}_j} \{\sum_k s_k \, g(t_k)\}^2}{\sum g^2(t_k)}}$$

The previous line and Eq. (5) match. Earlier, I showed Eq. (5) gives rise to Eq. (6) after we demean *t*, and that Eq. (6) and Eq. (4) are equivalent when $J = 1$ if we substitute the scaled Schoenfelds for the unscaled Schoenfelds.

Appendix B
Raw Results Output: Scenarios 1 and 2

NOTE: The figures corresponding to those in the main text (for $g(t) = \ln(t)$) are included in the supplemental materials. The supplemental materials also include a viewing app, written in R (`shiny`), to make viewing the relevant results easier, accessible here: https://bit.ly/3vjNe5B.

For descriptive information about the simulated datasets, see Appendix C.II.

TABLE 3. Mean PH Test *p*-values, Scenario 1

RC Pattern: Random (25% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.2132	0.5405	0.3975	0.4274
x_1^2		0.5949		0.4433
x_2	0.0199	0.0201	0.0028	0.0016
Global Test	0.0476	0.0799	0.0030	0.0031
$g(t) = \ln(t)$				
x_1	0.0864	0.5381	0.4722	0.5402
x_1^2		0.5217		0.5397
x_2	0.0003	0.0003	0.0004	0.0001
Global Test	0.0006	0.0013	0.0005	0.0007
$g(t) = \text{KM}$				
x_1	0.0915	0.5049	0.4530	0.4706
x_1^2		0.5446		0.4523
x_2	0.0004	0.0004	0.0006	0.0002
Global Test	0.0008	0.0016	0.0005	0.0006
$g(t) = \text{rank}$				
x_1	0.0911	0.5110	0.4614	0.4772
x_1^2		0.5448		0.4550
x_2	0.0003	0.0003	0.0005	0.0001
Global Test	0.0006	0.0012	0.0004	0.0005

RC Pattern: Largest 25%

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.1641	0.4588	0.4575	0.3890
x_1^2		0.4753		0.3331
x_2	0.0088	0.0077	0.0132	0.0054
Global Test	0.0141	0.0191	0.0138	0.0121
$g(t) = \ln(t)$				
x_1	0.1407	0.5462	0.5029	0.4825
x_1^2		0.5097		0.4820
x_2	0.0041	0.0046	0.0042	0.0017
Global Test	0.0080	0.0151	0.0058	0.0071
$g(t) = \text{KM}$				
x_1	0.1521	0.5154	0.4710	0.3650
x_1^2		0.5176		0.3300
x_2	0.0038	0.0037	0.0053	0.0017
Global Test	0.0065	0.0105	0.0058	0.0054
$g(t) = \text{rank}$				
x_1	0.1521	0.5154	0.4710	0.3649
x_1^2		0.5176		0.3299
x_2	0.0038	0.0037	0.0053	0.0017
Global Test	0.0065	0.0105	0.0058	0.0054

RC Pattern: N/A (0% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.1591	0.5296	0.3716	0.4134
x_1^2		0.5914		0.4323
x_2	0.0077	0.0077	0.0010	0.0004
Global Test	0.0205	0.0374	0.0008	0.0008
$g(t) = \ln(t)$				
x_1	0.0481	0.5298	0.4600	0.5438
x_1^2		0.5142		0.5468
x_2	0.0000	0.0001	0.0001	0.0000
Global Test	0.0001	0.0002	0.0001	0.0001
$g(t) = \text{KM}$				
x_1	0.0510	0.4985	0.4524	0.4694
x_1^2		0.5386		0.4386
x_2	0.0001	0.0001	0.0001	0.0000
Global Test	0.0001	0.0002	0.0001	0.0001
$g(t) = \text{rank}$				
x_1	0.0510	0.4985	0.4524	0.4694
x_1^2		0.5386		0.4385
x_2	0.0001	0.0001	0.0001	0.0000
Global Test	0.0001	0.0002	0.0001	0.0001

Notes

Reported quantity: mean of *p*-values from PH test for a covariate.

Shaded rows = should have average *p*-value less than 0.05.

"Wrg." = incorrect specification for base model (x_1^2 excluded), "Rgt." = correct specification for base model.

"Approx.": fast approximation of score test (`survival` 2.44-1), "Actual": actual score test (`survival` 3.2-11).

Reported in main text: $g(t) = \ln(t)$, largest 25%.

TABLE 4. Statistical Size/Power, Scenario 1

RC Pattern: Random (25% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.285	0.048	0.159	0.120
x_1^2		0.015		0.106
x_2	0.886	0.884	0.988	0.993
Global Test	0.769	0.667	0.986	0.985
$g(t) = \ln(t)$				
x_1	0.646	0.026	0.067	0.037
x_1^2		0.037		0.036
x_2	0.999	0.999	0.999	1
Global Test	0.998	0.994	0.999	0.998
$g(t) = \text{KM}$				
x_1	0.620	0.046	0.091	0.082
x_1^2		0.024		0.096
x_2	0.999	0.999	0.998	0.999
Global Test	0.997	0.993	0.998	0.998
$g(t) = \text{rank}$				
x_1	0.627	0.042	0.081	0.077
x_1^2		0.024		0.098
x_2	0.999	0.999	0.998	1
Global Test	0.998	0.995	0.999	0.999

RC Pattern: Largest 25%

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.493	0.089	0.085	0.170
x_1^2		0.073		0.243
x_2	0.963	0.966	0.943	0.978
Global Test	0.937	0.912	0.937	0.944
$g(t) = \ln(t)$				
x_1	0.517	0.022	0.047	0.092
x_1^2		0.039		0.082
x_2	0.983	0.980	0.986	0.995
Global Test	0.962	0.924	0.977	0.970
$g(t) = \text{KM}$				
x_1	0.510	0.042	0.070	0.200
x_1^2		0.039		0.225
x_2	0.984	0.984	0.980	0.994
Global Test	0.969	0.951	0.974	0.976
$g(t) = \text{rank}$				
x_1	0.509	0.042	0.070	0.200
x_1^2		0.039		0.225
x_2	0.984	0.984	0.980	0.994
Global Test	0.969	0.951	0.974	0.976

RC Pattern: N/A (0% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.373	0.055	0.188	0.137
x_1^2		0.017		0.119
x_2	0.960	0.961	0.996	0.999
Global Test	0.892	0.812	0.996	0.997
$g(t) = \ln(t)$				
x_1	0.783	0.031	0.074	0.033
x_1^2		0.043		0.034
x_2	1	1	1	1
Global Test	1	0.999	1	1
$g(t) = \text{KM}$				
x_1	0.767	0.050	0.090	0.085
x_1^2		0.028		0.110
x_2	1	1	1	1
Global Test	1	0.999	1	1
$g(t) = \text{rank}$				
x_1	0.767	0.050	0.090	0.085
x_1^2		0.028		0.110
x_2	1	1	1	1
Global Test	1	0.999	1	1

Notes

Reported quantity: proportion of *p*-values below 0.05 (0 = none below, 1 = all below).

Shaded rows: statistical power (≥ 80% = ideal).
Unshaded rows: statistical size (≤ 5%).

"Wrg." = incorrect specification for base model (x_1^2 excluded), "Rgt." = correct specification for base model.

"Approx.": fast approximation of score test (`survival` 2.44-1), "Actual": actual score test (`survival` 3.2-11).

TABLE 5. Mean PH Test *p*-values, Scenario 2

RC Pattern: Random (25% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.5620	0.4765	0.4721	0.4607
x_1x_2		0.5791		0.0553
x_2	0.3555	0.5547	0.0377	0.0360
Global Test	0.4679	0.5662	0.0542	0.0615
$g(t) = \ln(t)$				
x_1	0.5025	0.4285	0.4862	0.4643
x_1x_2		0.5045		0.0464
x_2	0.0525	0.3750	0.0260	0.0250
Global Test	0.0802	0.1080	0.0396	0.0476
$g(t) = \text{KM}$				
x_1	0.5107	0.4506	0.4875	0.4691
x_1x_2		0.5251		0.0600
x_2	0.0826	0.4046	0.0382	0.0365
Global Test	0.1223	0.1663	0.0564	0.0652
$g(t) = \text{rank}$				
x_1	0.5074	0.4510	0.4892	0.4704
x_1x_2		0.5189		0.0591
x_2	0.0675	0.3850	0.0366	0.0350
Global Test	0.1001	0.1358	0.0544	0.0634

RC Pattern: Largest 25%

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.5082	0.4472	0.5022	0.4892
x_1x_2		0.4864		0.0617
x_2	0.0394	0.3567	0.0368	0.0365
Global Test	0.0582	0.0739	0.0549	0.0668
$g(t) = \ln(t)$				
x_1	0.4902	0.4103	0.4948	0.4831
x_1x_2		0.4599		0.0406
x_2	0.0237	0.3436	0.0211	0.0211
Global Test	0.0359	0.0472	0.0329	0.0419
$g(t) = \text{KM}$				
x_1	0.5074	0.4524	0.5035	0.4911
x_1x_2		0.4876		0.0621
x_2	0.0384	0.3572	0.0360	0.0358
Global Test	0.0568	0.0722	0.0541	0.0663
$g(t) = \text{rank}$				
x_1	0.5074	0.4524	0.5035	0.4911
x_1x_2		0.4876		0.0621
x_2	0.0384	0.3572	0.0360	0.0358
Global Test	0.0568	0.0722	0.0541	0.0663

RC Pattern: N/A (0% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.5569	0.4685	0.4628	0.4597
x_1x_2		0.5709		0.0265
x_2	0.2921	0.5400	0.0155	0.0146
Global Test	0.4120	0.5165	0.0248	0.0297
$g(t) = \ln(t)$				
x_1	0.5028	0.4183	0.4841	0.4676
x_1x_2		0.4978		0.0215
x_2	0.0241	0.3388	0.0097	0.0093
Global Test	0.0405	0.0585	0.0162	0.0210
$g(t) = \text{KM}$				
x_1	0.5072	0.4406	0.4864	0.4773
x_1x_2		0.5137		0.0292
x_2	0.0332	0.3464	0.0151	0.0144
Global Test	0.0548	0.0799	0.0251	0.0311
$g(t) = \text{rank}$				
x_1	0.5072	0.4406	0.4864	0.4773
x_1x_2		0.5137		0.0292
x_2	0.0332	0.3464	0.0151	0.0144
Global Test	0.0548	0.0799	0.0251	0.0311

Notes

Reported quantity: mean of *p*-values from PH test for a covariate.

Shaded rows = should have average *p*-value less than 0.05.

"Wrg." = incorrect specification for base model (x_1x_2 excluded), "Rgt." = correct specification for base model.

"Approx.": fast approximation of score test (`survival` 2.44-1), "Actual": actual score test (`survival` 3.2-11).

Reported in main text: $g(t) = \ln(t)$, largest 25%.

TABLE 6. Statistical Size/Power, Scenario 2

RC Pattern: Random (25% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.014	0.065	0.068	0.077
x_1x_2		0.013		0.776
x_2	0.023	0.021	0.845	0.851
Global Test	0.014	0.022	0.780	0.750
$g(t) = \ln(t)$				
x_1	0.044	0.119	0.057	0.076
x_1x_2		0.050		0.812
x_2	0.768	0.154	0.892	0.895
Global Test	0.656	0.572	0.838	0.803
$g(t) = \text{KM}$				
x_1	0.039	0.099	0.060	0.073
x_1x_2		0.037		0.760
x_2	0.624	0.123	0.839	0.846
Global Test	0.483	0.388	0.765	0.738
$g(t) = \text{rank}$				
x_1	0.042	0.098	0.059	0.071
x_1x_2		0.042		0.763
x_2	0.695	0.141	0.846	0.852
Global Test	0.575	0.479	0.771	0.743

RC Pattern: Largest 25%

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.046	0.099	0.050	0.061
x_1x_2		0.061		0.759
x_2	0.830	0.187	0.842	0.844
Global Test	0.755	0.697	0.771	0.724
$g(t) = \ln(t)$				
x_1	0.057	0.139	0.050	0.061
x_1x_2		0.082		0.834
x_2	0.901	0.213	0.911	0.911
Global Test	0.845	0.797	0.860	0.822
$g(t) = \text{KM}$				
x_1	0.046	0.096	0.051	0.060
x_1x_2		0.060		0.756
x_2	0.835	0.187	0.844	0.847
Global Test	0.759	0.704	0.773	0.726
$g(t) = \text{rank}$				
x_1	0.046	0.096	0.051	0.060
x_1x_2		0.060		0.756
x_2	0.835	0.187	0.844	0.847
Global Test	0.759	0.704	0.773	0.726

RC Pattern: N/A (0% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.012	0.064	0.075	0.080
x_1x_2		0.013		0.880
x_2	0.027	0.022	0.932	0.937
Global Test	0.016	0.024	0.885	0.864
$g(t) = \ln(t)$				
x_1	0.045	0.126	0.061	0.073
x_1x_2		0.052		0.907
x_2	0.886	0.196	0.958	0.960
Global Test	0.806	0.730	0.925	0.905
$g(t) = \text{KM}$				
x_1	0.044	0.105	0.063	0.070
x_1x_2		0.044		0.868
x_2	0.832	0.181	0.931	0.936
Global Test	0.725	0.633	0.882	0.856
$g(t) = \text{rank}$				
x_1	0.044	0.105	0.063	0.070
x_1x_2		0.044		0.868
x_2	0.832	0.181	0.931	0.936
Global Test	0.725	0.633	0.882	0.856

Notes

Reported quantity: proportion of p-values below 0.05 (0 = none below, 1 = all below).

Shaded rows: statistical power (≥ 80% = ideal).
Unshaded rows: statistical size (≤ 5%).

"Wrg." = incorrect specification for base model (x_1x_2 excluded), "Rgt." = correct specification for base model.

"Approx.": fast approximation of score test (`survival` 2.44-1), "Actual": actual score test (`survival` 3.2-11).

Appendix C
x_1 Range, Magnitude

Keele's original x_1 distribution (uniform integers, [22,90]) affects the hazard's size appreciably, given his selected DGPs and parameter values (main text, fn. 4). The resulting hazards are large, meaning subjects fail almost immediately. The Cox model's performance is adversely affected in the presence of many ties (Hertz-Picciotto and Rockhill 1997)—when many subjects fail at the same time—and the issue appears to be particularly acute when many subjects fail in the (0,1] interval (Metzger and Jones 2022, Appendix I). As a consequence, it is possible that Keele's original findings are contingent on subjects failing very quickly.

I do two things to investigate this possibility. First, I run two additional scenarios in which I reduce the range and magnitude of x_1's values. Modifying x_1's properties reduces $h(t)$'s magnitude, in turn preventing `simsurv` from throwing an error, allowing me to use Keele's original parameter values. This also allows me to verify that my different x_1-related parameter values for Scenarios 1 and 2 are not somehow affecting the results I report, relative to Keele's. Second, I store descriptive statistics about the duration from every simulation draw for a given scenario–RC percentage–RC pattern triple, to see whether the triple has many subjects failing quickly. My rough rule-of-thumb definition for "quickly" is more than 50% of subjects failing in $t \in (0,1]$, motivated by Metzger and Jones' (2022, Appendix I) simulations investigating the Cox model's performance in the presence of ties and other work cited therein.

I. Simulations: Scenarios 3 and 4

I run two new simulation scenarios, summarized in Table 7. The x_1-related parameter values now reflect Keele's original choices, but x_1's distribution differs.

TABLE 7. Simulated Data: Setup Details, Scs. 3–4

	Keele (2010) [same as main text]	*Metzger Rerun*
n	100	100
$h_0(t)$		
Funct. Form.	exponential	exponential
Scale	0.15	0.15
Covariates		
x_1	Uniform integers, [22,90]	**Uniform, [0,1]**
x_2	Binomial with $p = 0.5$	Binomial with $p = 0.5$
True Linear Combo		
Sc. 3: Quadratic	$0.1x_1^2 + 1x_2 \ln(t)$	$\mathbf{0.1}x_1^2 + 1x_2 \ln(t)$
Sc. 4: Interactive	$0.1x_1 + 1x_2 \ln(t) + 0.4x_1x_2$	$\mathbf{0.1}x_1 + 1x_2 \ln(t) + \mathbf{0.4}x_1x_2$
Right Censoring		
% RC	25% [6]	25%
RC Type	(see fn. 6)	{Random, Largest *rc*%}

NOTE: tan shading and/or bold parameter values = difference from main text's Scs. 1 and 2

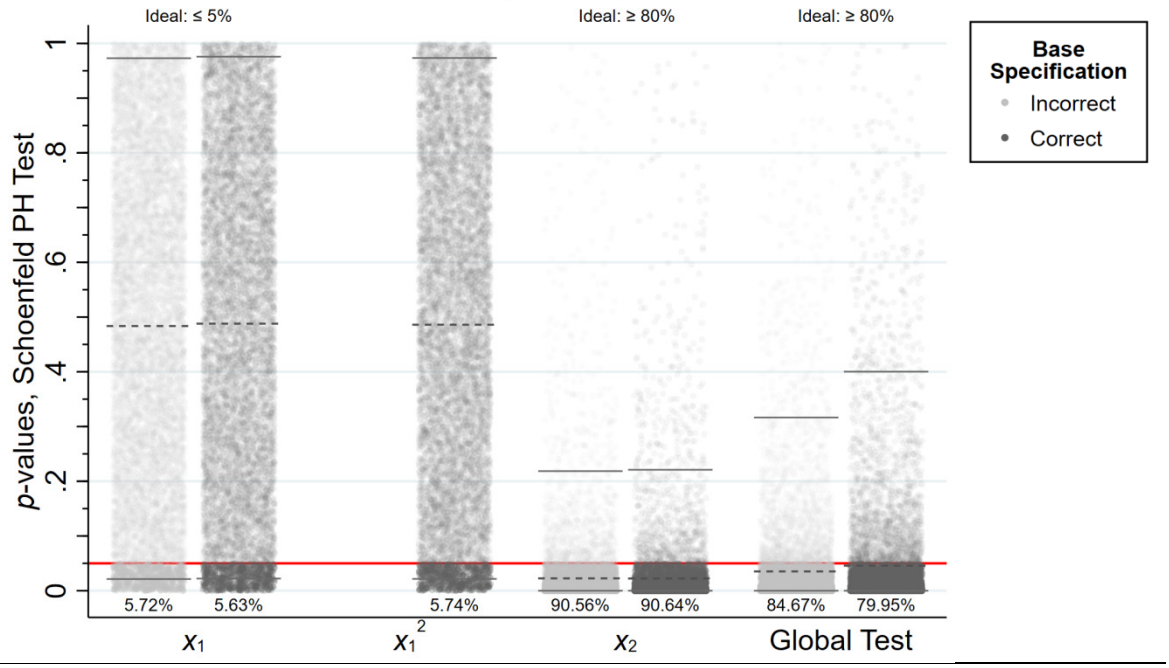
A. Results: Graphs

Figures 3 and 4 (below) are comparable to the main text's Figures 1 and 2: they report the results for $g(t) = \ln(t)$, with the largest 25% of durations right censored. The approximated PH test appears in panel (a); the actual PH test, in panel (b). The full set of results, for all $g(t)$s, RC patterns, and RC percentages, are reported in the next subsection (Appendix C.I.B).

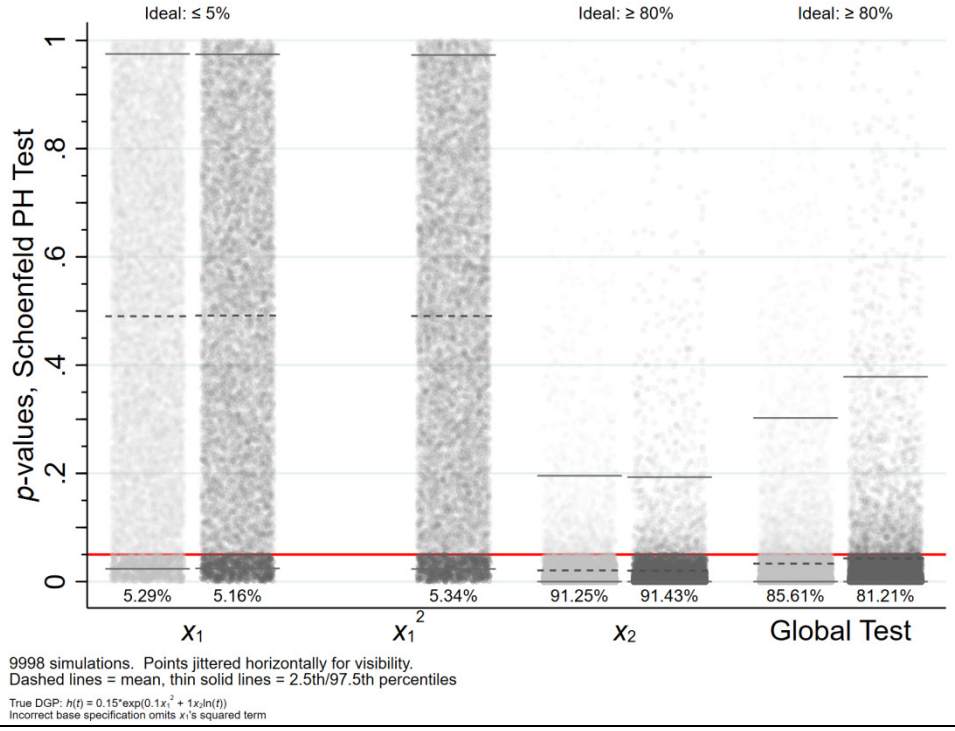
As I mention in the main text (Section III.C), I find *no* evidence of performance issues due to misspecification in these scenarios, regardless of which version of the PH test calculation I use, the RC amount, or the RC pattern type. This pattern is particularly illuminating for Scenario 3, which is analogous to Scenario 1. Scenario 1 was the only scenario whose results for the approximated PH test were somewhat similar to Keele's—a pattern that does not continue in Scenario 3. However, Scenario 1 is the only scenario in which > 50% of subjects fail in $t \in (0,1]$ (this appendix, next subsection [Table 12]). This evidence is far from conclusive, in terms of scope conditions affecting the approximated PH test's behavior, but it is suggestive. Probing these scope conditions more deeply is left to future research.

FIGURE 3. Scenario 3 Simulations (n = 100)
(DGP = akin to Scenario 1)

(a) PH Test: Approximation



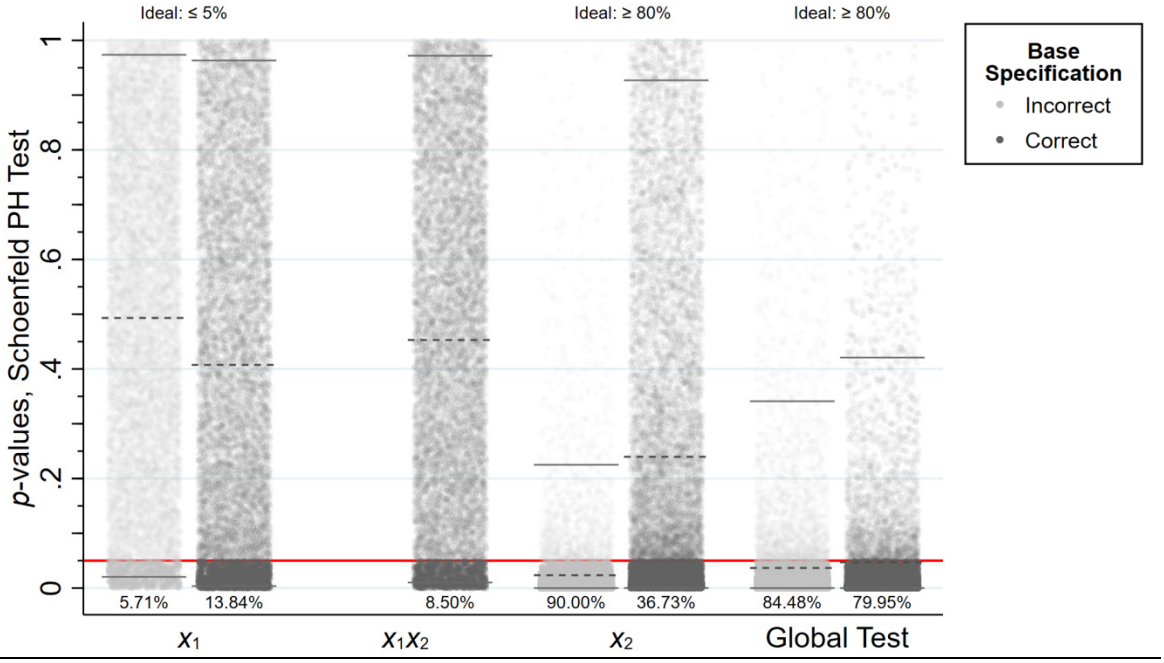
(b) PH Test: Actual



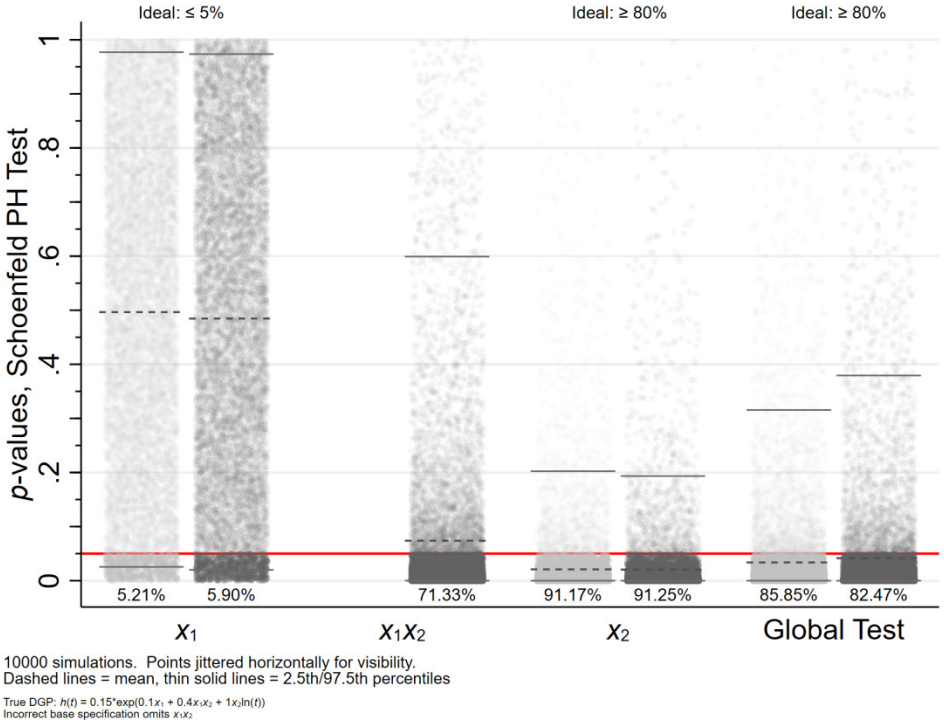
Note: RC pattern = top 25% of durations are right censored. $g(t) = \ln(t)$ reported.

FIGURE 4. Scenario 4 Simulations (n = 100)
(DGP = akin to Scenario 2)

(a) PH Test: Approximation



(b) PH Test: Actual



Note: RC pattern = top 25% of durations are right censored. $g(t) = \ln(t)$ reported.

B. Results: Raw Output (All)

TABLE 8. Mean PH Test *p*-values, Scenario 3

RC Pattern: Random (25% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.5677	0.5659	0.4764	0.4739
x_1^2		0.5686		0.4727
x_2	0.3008	0.3037	0.0330	0.0319
Global Test	0.4201	0.4940	0.0489	0.0573
$g(t) = \ln(t)$				
x_1	0.5033	0.5036	0.4895	0.4890
x_1^2		0.5015		0.4890
x_2	0.0425	0.0431	0.0223	0.0218
Global Test	0.0664	0.0864	0.0347	0.0437
$g(t) = \text{KM}$				
x_1	0.5115	0.5125	0.4916	0.4902
x_1^2		0.5141		0.4900
x_2	0.0697	0.0707	0.0338	0.0328
Global Test	0.1056	0.1351	0.0502	0.0616
$g(t) = \text{rank}$				
x_1	0.5076	0.5087	0.4927	0.4917
x_1^2		0.5093		0.4910
x_2	0.0569	0.0577	0.0323	0.0315
Global Test	0.0858	0.1102	0.0481	0.0594

RC Pattern: Largest 25%

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.5015	0.5046	0.4951	0.4962
x_1^2		0.5048		0.4953
x_2	0.0379	0.0374	0.0362	0.0350
Global Test	0.0551	0.0698	0.0529	0.0655
$g(t) = \ln(t)$				
x_1	0.4836	0.4879	0.4903	0.4915
x_1^2		0.4860		0.4907
x_2	0.0226	0.0224	0.0207	0.0202
Global Test	0.0353	0.0457	0.0332	0.0429
$g(t) = \text{KM}$				
x_1	0.5009	0.5027	0.4952	0.4962
x_1^2		0.5025		0.4949
x_2	0.0379	0.0375	0.0363	0.0352
Global Test	0.0549	0.0693	0.0530	0.0657
$g(t) = \text{rank}$				
x_1	0.5009	0.5027	0.4952	0.4962
x_1^2		0.5025		0.4949
x_2	0.0379	0.0375	0.0363	0.0352
Global Test	0.0549	0.0693	0.0530	0.0657

RC Pattern: N/A (0% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.5673	0.5681	0.4725	0.4692
x_1^2		0.5691		0.4691
x_2	0.2368	0.2397	0.0127	0.0121
Global Test	0.3582	0.4410	0.0214	0.0261
$g(t) = \ln(t)$				
x_1	0.5022	0.4993	0.4860	0.4872
x_1^2		0.4979		0.4875
x_2	0.0183	0.0187	0.0077	0.0075
Global Test	0.0323	0.0447	0.0140	0.0186
$g(t) = \text{KM}$				
x_1	0.5096	0.5114	0.4919	0.4912
x_1^2		0.5096		0.4898
x_2	0.0264	0.0271	0.0127	0.0122
Global Test	0.0452	0.0618	0.0219	0.0285
$g(t) = \text{rank}$				
x_1	0.5096	0.5114	0.4919	0.4912
x_1^2		0.5096		0.4898
x_2	0.0264	0.0271	0.0127	0.0122
Global Test	0.0452	0.0618	0.0219	0.0285

Notes

Reported quantity: mean of *p*-values from PH test for a covariate.

Shaded rows = should have average *p*-value less than 0.05.

"Wrg." = incorrect specification for base model (x_1^2 excluded), "Rgt." = correct specification for base model.

"Approx.": fast approximation of score test (`survival` 2.44-1), "Actual": actual score test (`survival` 3.2-11).

Reported in main text: $g(t) = \ln(t)$, largest 25%.

TABLE 9. Statistical Size/Power, Scenario 3

RC Pattern: Random (25% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.012	0.012	0.060	0.061
x_1^2		0.012		0.062
x_2	0.049	0.052	0.859	0.862
Global Test	0.027	0.020	0.796	0.766
$g(t) = \ln(t)$				
x_1	0.046	0.045	0.058	0.057
x_1^2		0.046		0.058
x_2	0.811	0.809	0.906	0.908
Global Test	0.706	0.630	0.852	0.809
$g(t) =$ KM				
x_1	0.042	0.040	0.056	0.057
x_1^2		0.042		0.061
x_2	0.678	0.677	0.858	0.863
Global Test	0.539	0.454	0.790	0.743
$g(t) =$ rank				
x_1	0.045	0.043	0.056	0.056
x_1^2		0.044		0.059
x_2	0.745	0.743	0.866	0.872
Global Test	0.622	0.541	0.794	0.750

RC Pattern: Largest 25%

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.049	0.046	0.053	0.052
x_1^2		0.047		0.052
x_2	0.841	0.844	0.847	0.851
Global Test	0.764	0.711	0.775	0.732
$g(t) = \ln(t)$				
x_1	0.057	0.056	0.053	0.052
x_1^2		0.057		0.053
x_2	0.906	0.906	0.912	0.914
Global Test	0.847	0.799	0.856	0.812
$g(t) =$ KM				
x_1	0.048	0.048	0.054	0.052
x_1^2		0.048		0.052
x_2	0.843	0.845	0.848	0.854
Global Test	0.767	0.713	0.777	0.728
$g(t) =$ rank				
x_1	0.048	0.048	0.054	0.052
x_1^2		0.048		0.052
x_2	0.843	0.845	0.848	0.854
Global Test	0.767	0.713	0.777	0.728

RC Pattern: N/A (0% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.008	0.011	0.063	0.067
x_1^2		0.010		0.070
x_2	0.067	0.068	0.940	0.943
Global Test	0.027	0.018	0.901	0.878
$g(t) = \ln(t)$				
x_1	0.044	0.047	0.057	0.058
x_1^2		0.048		0.058
x_2	0.914	0.912	0.965	0.966
Global Test	0.845	0.787	0.936	0.913
$g(t) =$ KM				
x_1	0.041	0.046	0.056	0.057
x_1^2		0.047		0.058
x_2	0.868	0.866	0.941	0.944
Global Test	0.777	0.703	0.900	0.866
$g(t) =$ rank				
x_1	0.041	0.046	0.056	0.057
x_1^2		0.047		0.058
x_2	0.868	0.866	0.941	0.944
Global Test	0.777	0.703	0.900	0.866

Notes

Reported quantity: proportion of p-values below 0.05 (0 = none below, 1 = all below).

Shaded rows: statistical power (≥ 80% = ideal).
Unshaded rows: statistical size (≤ 5%).

"Wrg." = incorrect specification for base model (x_1^2 excluded), "Rgt." = correct specification for base model.

"Approx.": fast approximation of score test (`survival` 2.44-1), "Actual": actual score test (`survival` 3.2-11).

TABLE 10. Mean PH Test *p*-values, Scenario 4

RC Pattern: Random (25% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.5583	0.4756	0.4643	0.4590
x_1x_2		0.5769		0.0924
x_2	0.3529	0.5206	0.0390	0.0364
Global Test	0.4602	0.5572	0.0560	0.0626
$g(t) = \ln(t)$				
x_1	0.5002	0.4242	0.4789	0.4582
x_1x_2		0.5013		0.0843
x_2	0.0527	0.2632	0.0259	0.0246
Global Test	0.0785	0.1049	0.0394	0.0471
$g(t) = \text{KM}$				
x_1	0.5053	0.4453	0.4748	0.4648
x_1x_2		0.5218		0.1014
x_2	0.0847	0.2990	0.0392	0.0367
Global Test	0.1220	0.1630	0.0581	0.0652
$g(t) = \text{rank}$				
x_1	0.5023	0.4453	0.4768	0.4655
x_1x_2		0.5157		0.1012
x_2	0.0695	0.2751	0.0374	0.0352
Global Test	0.1004	0.1333	0.0561	0.0633

RC Pattern: Largest 25%

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.5124	0.4473	0.4971	0.4878
x_1x_2		0.4796		0.1019
x_2	0.0407	0.2525	0.0377	0.0365
Global Test	0.0605	0.0740	0.0571	0.0666
$g(t) = \ln(t)$				
x_1	0.4932	0.4075	0.4966	0.4847
x_1x_2		0.4529		0.0739
x_2	0.0235	0.2396	0.0209	0.0205
Global Test	0.0368	0.0471	0.0338	0.0416
$g(t) = \text{KM}$				
x_1	0.5097	0.4520	0.4982	0.4882
x_1x_2		0.4815		0.1034
x_2	0.0394	0.2515	0.0367	0.0355
Global Test	0.0588	0.0724	0.0561	0.0663
$g(t) = \text{rank}$				
x_1	0.5097	0.4520	0.4982	0.4882
x_1x_2		0.4815		0.1034
x_2	0.0394	0.2515	0.0367	0.0355
Global Test	0.0588	0.0724	0.0561	0.0663

RC Pattern: N/A (0% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.5552	0.4723	0.4573	0.4588
x_1x_2		0.5765		0.0517
x_2	0.2884	0.4939	0.0162	0.0146
Global Test	0.4023	0.5079	0.0260	0.0304
$g(t) = \ln(t)$				
x_1	0.4985	0.4141	0.4744	0.4602
x_1x_2		0.4978		0.0463
x_2	0.0241	0.2072	0.0096	0.0089
Global Test	0.0393	0.0566	0.0162	0.0209
$g(t) = \text{KM}$				
x_1	0.4989	0.4363	0.4698	0.4668
x_1x_2		0.5133		0.0588
x_2	0.0345	0.2187	0.0155	0.0141
Global Test	0.0546	0.0780	0.0260	0.0311
$g(t) = \text{rank}$				
x_1	0.4989	0.4363	0.4698	0.4668
x_1x_2		0.5133		0.0588
x_2	0.0345	0.2187	0.0155	0.0141
Global Test	0.0546	0.0780	0.0260	0.0311

Notes

Reported quantity: mean of *p*-values from PH test for a covariate.

Shaded rows = should have average *p*-value less than 0.05.

"Wrg." = incorrect specification for base model (x_1x_2 excluded), "Rgt." = correct specification for base model.

"Approx.": fast approximation of score test (`survival` 2.44-1), "Actual": actual score test (`survival` 3.2-11).

Reported in main text: $g(t) = \ln(t)$, largest 25%.

TABLE 11. Statistical Size/Power, Scenario 4

RC Pattern: Random (25% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.016	0.063	0.072	0.077
x_1x_2		0.013		0.659
x_2	0.025	0.031	0.834	0.845
Global Test	0.016	0.024	0.771	0.748
$g(t) = \ln(t)$				
x_1	0.044	0.121	0.061	0.079
x_1x_2		0.050		0.681
x_2	0.765	0.283	0.891	0.897
Global Test	0.670	0.585	0.834	0.804
$g(t) =$ KM				
x_1	0.041	0.098	0.065	0.080
x_1x_2		0.039		0.631
x_2	0.622	0.226	0.832	0.841
Global Test	0.491	0.399	0.762	0.741
$g(t) =$ rank				
x_1	0.044	0.098	0.064	0.080
x_1x_2		0.043		0.632
x_2	0.692	0.267	0.839	0.845
Global Test	0.580	0.486	0.769	0.747

RC Pattern: Largest 25%

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.047	0.099	0.051	0.059
x_1x_2		0.063		0.640
x_2	0.823	0.334	0.837	0.840
Global Test	0.749	0.701	0.763	0.730
$g(t) = \ln(t)$				
x_1	0.057	0.138	0.052	0.059
x_1x_2		0.085		0.713
x_2	0.900	0.367	0.912	0.913
Global Test	0.845	0.799	0.859	0.825
$g(t) =$ KM				
x_1	0.047	0.097	0.052	0.059
x_1x_2		0.060		0.636
x_2	0.830	0.337	0.841	0.845
Global Test	0.756	0.706	0.768	0.731
$g(t) =$ rank				
x_1	0.047	0.097	0.052	0.059
x_1x_2		0.060		0.636
x_2	0.830	0.337	0.841	0.845
Global Test	0.756	0.706	0.768	0.731

RC Pattern: N/A (0% RC)

	PH TEST: APPROX.		PH TEST: ACTUAL	
	Wrg.	*Rgt.*	*Wrg.*	*Rgt.*
$g(t) = t$				
x_1	0.013	0.065	0.080	0.082
x_1x_2		0.012		0.775
x_2	0.033	0.036	0.925	0.932
Global Test	0.017	0.026	0.879	0.861
$g(t) = \ln(t)$				
x_1	0.047	0.127	0.064	0.076
x_1x_2		0.051		0.802
x_2	0.886	0.373	0.960	0.962
Global Test	0.814	0.742	0.926	0.907
$g(t) =$ KM				
x_1	0.045	0.102	0.071	0.073
x_1x_2		0.042		0.752
x_2	0.826	0.340	0.927	0.934
Global Test	0.735	0.650	0.878	0.857
$g(t) =$ rank				
x_1	0.045	0.102	0.071	0.073
x_1x_2		0.042		0.752
x_2	0.826	0.340	0.927	0.934
Global Test	0.735	0.650	0.878	0.857

Notes

Reported quantity: proportion of p-values below 0.05 (0 = none below, 1 = all below).

Shaded rows: statistical power (≥ 80% = ideal).
Unshaded rows: statistical size (≤ 5%).

"Wrg." = incorrect specification for base model (x_1x_2 excluded), "Rgt." = correct specification for base model.

"Approx.": fast approximation of score test (`survival` 2.44-1), "Actual": actual score test (`survival` 3.2-11).

II. Descriptive Statistics

TABLE 12. Descriptive Statistics, All Scenarios

RC Pattern: Random (25% RC)

Scen.	*nEvents*	*KM_75th*	*KM_50th*	*KM_25th*
1	74.9969	0.1233	0.5955	1.8214
2	75.0388	2.1091	3.7983	7.5302
3	75.0342	2.3141	4.1508	7.8239
4	75.0293	2.1292	3.8369	7.5585

RC Pattern: Largest 25%

Scen.	*nEvents*	*KM_75th*	*KM_50th*	*KM_25th*
1	75	0.0712	0.3395	1.1438
2	75	1.7199	3.0733	5.2552
3	75	1.8822	3.3725	5.6364
4	75	1.7348	3.1084	5.3022

RC Pattern: N/A (0% RC)

Scen.	*nEvents*	*KM_75th*	*KM_50th*	*KM_25th*
1	100	0.0712	0.3395	1.1438
2	100	1.7199	3.0733	5.2552
3	100	1.8822	3.3725	5.6364
4	100	1.7348	3.1084	5.3022

Legend:

- `nEvents`: the number of observed failure events in a given draw, averaged across all simulation draws
- `KM_75th`: the *t* satisfying $S_{KM}(t) = 0.75$ in a given draw, averaged across all simulation draws; 75% of a draw's subjects fail after this *t* value.
- `KM_50th`: the median duration—the *t* satisfying $S_{KM}(t) = 0.5$ in a given draw, averaged across all simulation draws; half of a draw's subjects fail after this *t* value.
- `KM_25th`: the *t* satisfying $S_{KM}(t) = 0.25$ in a given draw, averaged across all simulation draws; 25% of a draw's subjects fail after this *t* value.

Appendix D
Output from Keele's Original Replication File

(starts next page)

June 20, 2021

The results below are generated from an R script.

```
# SKM tweaks flagged with "SKM"; else, original code

library(survival, lib.loc="                    Keele redux/data/_survPkgVers/2.18")

## Loading required package:  splines

# ^ to load version available in mid-2005 @ time of Keele's writing [SKM]

set.seed(316857)

n <- 100
caseid <- (1:n)
B.age <- .1
B.treat <- 1
rxe <- 1.0
prx <- 0.5
duration <- 2
base <- 0.15
pcensor <- 0.25
rx   <- rbinom(n, 1, prx)       #Treatment variable Gen from Binomial Dist.
age <- round(runif(n, 22, 90))
age.sq <- age^2

hazard <- base * exp(B.age*age^2 + B.treat*(rx*log(caseid)))
    # get hazard min/max, for tShoot purposes [SKM]
    range(hazard)

## [1] 1.342743e+24          Inf

time <- rexp(n) / hazard

# censoring set up
censor <- rbinom(n, 1, pcensor)
if (duration > 0){
    test <- (time <= duration)
} else {
    test <- rep(1, n)
}
status <- test + (1 - censor)
status <- (status > 0)
time <- time * test + duration * (1 - test)
  # take peek at naive descriptives for time [SKM]
  summary(time)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.000e+00 0.000e+00 0.000e+00 6.659e-27 0.000e+00 5.646e-25

data.1 <- data.frame(caseid, time, status, rx, age, age.sq, hazard)

#Linear Form
mod.1 <- coxph(Surv(time, status) ~ age + rx, data = data.1)
    # check status var, since he uses this as his fail var (should = ~75) [SKM]
    sum(status)

## [1] 100

mod.1

## Call:
## coxph(formula = Surv(time, status) ~ age + rx, data = data.1)
##
##
##      coef exp(coef) se(coef)    z      p
## age 1.601      4.96    0.186 8.60 0.0000
## rx  0.847      2.33    0.272 3.12 0.0018
##
## Likelihood ratio test=524  on 2 df, p=0  n= 100

cox.zph(mod.1)

##          rho chisq        p
## age    0.393 49.86 1.65e-12
## rx     0.222  4.17 4.12e-02
## GLOBAL    NA 50.10 1.32e-11

#Nonlinearity
mod.2 <- coxph(Surv(time, status) ~ pspline(age, df=4) + rx, data = data.1)
cox.zph(mod.2)

##               rho  chisq     p
## ps(age)2  -0.1062 1.2120 0.271
## ps(age)3  -0.0938 1.2542 0.263
## ps(age)4  -0.0713 0.6454 0.422
## ps(age)5  -0.0537 0.2800 0.597
## ps(age)6  -0.0411 0.1218 0.727
## ps(age)7  -0.0342 0.0641 0.800
## ps(age)8  -0.0332 0.0485 0.826
## ps(age)9  -0.0377 0.0536 0.817
## ps(age)10 -0.0473 0.0778 0.780
## ps(age)11 -0.0577 0.1148 0.735
## ps(age)12 -0.0514 0.0914 0.762
## ps(age)13 -0.0606 0.1246 0.724
## rx         0.1390 1.6825 0.195
## GLOBAL         NA 4.4030 0.986

#Omitted Variable
mod.3 <- coxph(Surv(time, status) ~ rx, data = data.1)
mod.3
```

```
## Call:
## coxph(formula = Surv(time, status) ~ rx, data = data.1)
##
##
##       coef exp(coef) se(coef)     z    p
## rx 0.0216      1.02    0.203 0.106 0.92
##
## Likelihood ratio test=0.01  on 1 df, p=0.915  n= 100
```

```
cox.zph(mod.3)
```

```
##        rho chisq     p
## rx -0.0708 0.492 0.483
```

```
################################################
# Interaction
################################################
B.int <- .4
#hazard <- base * exp(B.age*age + B.treat*(rx*log(case.id)) + B.int*(rx*age)) # ORIGINAL
hazard <- base * exp(B.age*age + B.treat*(rx*log(caseid)) + B.int*(rx*age))   # SKM: needs to switch to
    # get hazard min/max, for tShoot purposes [SKM]
    range(hazard)
```

```
## [1] 1.653476e+00 1.017058e+20
```

```
time <- rexp(n) / hazard

# censoring set up
censor <- rbinom(n, 1, pcensor)
if (duration > 0){
    test <- (time <= duration)
} else {
    test <- rep(1, n)
}
status <- test + (1 - censor)
status <- (status > 0)
time <- time * test + duration * (1 - test)
  # take peek at naive descriptives for time [SKM]
  summary(time)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000000 0.0000002 0.0418283 0.0116288 0.6626765
```

```
data.2 <- data.frame(caseid, time, status, rx, age, age.sq, hazard)

#Omitted Interaction
mod.1 <- coxph(Surv(time, status) ~ age + rx, data = data.2)
    # check status var, since he uses this as his fail var (should = ~75) [SKM]
    sum(status)
```

```
## [1] 100
```

```
mod.1
```

```
## Call:
## coxph(formula = Surv(time, status) ~ age + rx, data = data.2)
```

```
##
##
##       coef exp(coef) se(coef)    z p
## age   0.16  1.17e+00   0.0163 9.82 0
## rx   13.28  5.87e+05   1.4642 9.07 0
##
## Likelihood ratio test=317  on 2 df, p=0  n= 100
```

```
cox.zph(mod.1)
```

```
##           rho chisq        p
## age    -0.396  22.7 1.87e-06
## rx     -0.294  10.2 1.42e-03
## GLOBAL     NA  22.7 1.16e-05
```

```
#Correct Specification
mod.2 <- coxph(Surv(time, status) ~ age + rx + age:rx, data = data.2)
mod.2
```

```
## Call:
## coxph(formula = Surv(time, status) ~ age + rx + age:rx, data = data.2)
##
##
##          coef exp(coef) se(coef)    z       p
## age    0.101      1.11   0.0157 6.41 1.4e-10
## rx     3.100     22.19   2.1701 1.43 1.5e-01
## age:rx 0.258      1.29   0.0520 4.96 7.0e-07
##
## Likelihood ratio test=354  on 3 df, p=0  n= 100
```

```
cox.zph(mod.2)
```

```
##             rho   chisq     p
## age     0.00216 0.00045 0.983
## rx     -0.01584 0.02468 0.875
## age:rx  0.02643 0.06269 0.802
## GLOBAL       NA 0.07343 0.995
```

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()
```

```
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19043)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252  LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252 LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] splines   stats     graphics  grDevices utils     datasets  methods   base
```

```
##
## other attached packages:
## [1] survival_2.18
##
## loaded via a namespace (and not attached):
##  [1] compiler_4.1.0 magrittr_2.0.1 pbdZMQ_0.3-5   tools_4.1.0    stringi_1.6.2
##  [6] highr_0.9      knitr_1.33     stringr_1.4.0  xfun_0.24      evaluate_0.14
```

```
Sys.time()
```

```
## [1] "2021-06-20 11:12:05 EDT"
```

Appendices Works Cited

Greene, W. H. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Hertz-Picciotto, I., and B. Rockhill. 1997. "Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression." *Biometrics* 53 (3): 1151–56.

Metzger, S. K., and B. T. Jones. 2022. "Getting Time Right: Using Cox Models and Probabilities to Interpret Binary Panel Data." *Political Analysis* 30 (2): 151–66.

Therneau, T. M. 2021. "`cox.zph`: `zph.rnw` Documentation." https://github.com/therneau/survival/blob/f2567b77252ac7935eba0ead364665c654ef28d3/noweb/zph.Rnw.

Therneau, T. M., and P. M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer.