Deriving an Analytical Solution to Inversion of Royston/Parmar Restricted Cubic Spline Parametric Survival Models for Discrete Event Simulation

Evidera | PPD

MSR146

Bungey, George[1]; Moller, Jorgen[2]

[1]Evidera Ltd., London, UK; [2]Evidera Ltd., Lund, Sweden

Background

- Royston/Parmar restricted cubic spline (RCS) models are more flexible extensions of Weibull, log-normal, and log-logistic parametric survival models that allow modeling of the hazard function as a restricted cubic spline instead of a linear function of log time.[1]
- These models are highlighted in National Institute for Health and Care Excellence (NICE) Decision Support Unit (DSU) Technical Support Document guidelines[2,3] as potential alternatives to standard parametric survival models, with other recently published guidance[4] also suggesting Royston/Parmar spline models may be preferable to other types of (non-cure) flexible parametric survival models (landmark, piecewise, parametric mixture).
- Discrete event simulation (DES) models simulate times to events rather than using cumulative survival probabilities from parametric survival models, with inversion of the survival functions required for analytical solutions to derive these event times from given survival estimates or sampling using random numbers.
- While numerical methods can be used to approximate event times for more complex survival models, this process may be slow, especially when repeated over large numbers of simulations.

Objectives

- We aimed to derive analytical solutions to inverse functions for Royston/Parmar RCS parametric survival models.

Methods

Royston/Parmar Spline Function

- For a time value t and x = ln(t), Royston/Parmar RCS cumulative survival functions (excluding covariates) with n internal knots, gamma parameters $\gamma_0, \dots, \gamma_{n+1}$ and knot values $k_{min}, k_1, \dots, k_n, k_{max}$ in log time ordered in increasing value can be described in the form:

$$S(t) = F(s(x,\gamma)) \quad \text{(Equation 1)}$$

Where F(z) for z = $s(x,\gamma)$ is defined as follows depending on the link function:

Cumulative Hazards: $F(z) = \exp(-\exp(z))$

Cumulative Odds: $F(z) = \frac{1}{(1+\exp(z))}$

Probit/Normal deviate: $F(z) = 1 - \Phi(z)$

Where Φ is the standard normal distribution, and s(x,γ) is a cubic polynomial function:

$$s(x,\gamma) = \gamma_0 + \gamma_1 x + \sum_{q=2}^{n+1} \gamma_q v_{q-1}(x) \quad \text{(Equation 2)}$$

With basis function $v_j(x)$ defined as:

$$v_j(x) = (x-k_j)_+^3 - \lambda_j(x-k_{min})_+^3 - (1-\lambda_j)(x-k_{max})_+^3,$$

Where $\lambda_j = \frac{k_{max}-k_j}{k_{max}-k_{min}}$ and $(x-p)_+ = \max(0, x-p)$.

Objective Function

- To invert the function S(t) to find a solution for t from a given cumulative survival estimate $\hat{S}$, we first invert F(z) for each link function as follows

Cumulative Hazards: $F^{-1}(z) = \ln(-\ln(z))$

Cumulative Odds: $F^{-1}(z) = \ln(\frac{1}{z} - 1)$

Probit/Normal deviate: $F^{-1}(z) = \Phi^{-1}(1-z)$

Where Φ^{-1} is the inverse standard normal distribution.

- By applying $F^{-1}(z)$ to both sides of Equation 1 and rearranging, this then allows us to describe our objective function to solve for x = ln(t) as:

$$s(x,\gamma) - F^{-1}(\hat{S}) = 0 \quad \text{(Equation 3)}$$

Defining Case Types

- As Royston/Parmar splines include linearity constraints before the first knot (k_{min}) and after the last knot (k_{max}), three main case types can be defined according to the positioning of the time estimate in relation to these boundary knot values.
- As we do not know the value of t, which we are trying to estimate from a given cumulative survival estimate $\hat{S}$, we instead define the case types according to the cumulative survival estimates produced by the boundary knots, given S(t) is a non-increasing function with increasing values of t:

Case 1: $\hat{S} \geq S(\exp(k_{min}))$

Case 2: $S(\exp(k_{min})) > \hat{S} \geq S(\exp(k_{max}))$

Case 3: $\hat{S} < S(\exp(k_{max}))$

Case 1 Solution

- For Case 1, we have $\hat{S} \geq S(\exp(k_{min}))$ which implies that $x \leq k_{min}$. As noted above, Royston/Parmar RCS models include linearity constraints before the first knot and after last knot; in this case the term $\sum_{q=2}^{n+1} \gamma_q v_{q-1}(\hat{x})$ in Equation 2 becomes 0, which means Equation 3 becomes:

$$\gamma_1 x + \gamma_0 - F^{-1}(\hat{S}) = 0 \quad \text{(Equation 4)}$$

- We now have a linear function of the form $cx + d = 0$, where c = γ_1 and $d = \gamma_0 - F^{-1}(\hat{S})$. Our solution for $t = \exp(x) = \exp(\frac{-d}{c})$ is therefore derived as:

$$t = exp(\frac{F^{-1}(\hat{S}) - \gamma_0}{\gamma_1})$$

Case 2 Solution

- For Case 2, we know that $S(\exp(k_{min})) > \hat{S} \geq S(\exp(k_{max}))$, which means that unlike Case 1 and Case 3, our cubic and quadratic terms do not cancel out, and we need to solve a cubic polynomial function. For any cubic equation $ax^3 + bx^2 + cx + d$, the three solutions for x (x_0, x_1, x_2) can be described in a generalized form[5] as:

$$x_m = -\frac{1}{3a}\left(b + \xi^m C + \frac{\Delta_0}{\xi^m C}\right), m \in \{0,1,2\} \quad \text{(Equation 5)}$$

Where

$$C = \sqrt[3]{\frac{\Delta_1 \pm \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}}$$

$$\Delta_0 = b^2 - 3ac$$

$$\Delta_1 = 2b^3 - 9abc + 27a^2 d$$

$$\xi = -\frac{1}{2} + \frac{\sqrt{3}}{2} i$$

With i being the imaginary number equivalent to $\sqrt{-1}$.

- As use of the positive or negative square root of $\Delta_1^2 - 4\Delta_0^3$ is mostly arbitrary[5] (except in the specific case C=0, handled separately below), we adopt the positive square root and redefine C as:

$$C = \sqrt[3]{\frac{\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}}$$

- From expanding the $v_j(x)$ basis function terms and combining with corresponding γ parameter terms, we can generalize the cubic, quadratic, linear, and constant terms as:

$$\sum_{q=2}^{n+1} \gamma_q v_{q-1}(x), x^3 \text{ terms: } \sum_{q=2}^{n+1} \gamma_q (1 - \lambda_{q-1}) = \kappa_3$$

$$\sum_{q=2}^{n+1} \gamma_q v_{q-1}(x), x^2 \text{ terms: } -3\sum_{q=2}^{n+1} \gamma_q (k_{q-1} - \lambda_{q-1} k_{min}) = -\kappa_2$$

$$\sum_{q=2}^{n+1} \gamma_q v_{q-1}(x), x \text{ terms: } 3\sum_{q=2}^{n+1} \gamma_q (k_{q-1}^2 - \lambda_{q-1} k_{min}^2) = \kappa_1$$

$$\sum_{q=2}^{n+1} \gamma_q v_{q-1}(x), \text{constant terms: } -\sum_{q=2}^{n+1} \gamma_q (k_{q-1}^3 - \lambda_{q-1} k_{min}^3) = -\kappa_0$$

- Combining these with the remaining terms in Equation 3, the coefficients of our cubic equation $ax^3 + bx^2 + cx + d$ we need to solve for x=ln(t) become:

$$a = \kappa_3$$

$$b = -\kappa_2$$

$$c = \gamma_1 + \kappa_1$$

$$d = \gamma_0 - F^{-1}(\hat{S}) - \kappa_0$$

Methods (cont.)

- Using the values for a, b, c, and d above we have values for Δ_0 and Δ_1 as:

$$\Delta_0 = \kappa_2^2 - 3\kappa_3(\gamma_1 + \kappa_1)$$

$$\Delta_1 = -2\kappa_2^3 + 9\kappa_2\kappa_3(\gamma_1 + \kappa_1) - 27\kappa_3^2(F^{-1}(\hat{S}) - \gamma_0 + \kappa_0)$$

- However, for calculating C, the term $\Delta_1^2 - 4\Delta_0^3$ within the square root can take positive and negative values, with negative values resulting in C^3 being a complex number. We further subdivide Case 2 into two sub-case types: Case 2a where C^3 is real ($\Delta_1^2 \geq 4\Delta_0^3$) and Case 2b where C^3 is complex ($\Delta_1^2 < 4\Delta_0^3$).

Case 2a

- For Case 2a, the term within the square root is positive. We then have a real cube root for C, which when incorporated into Equation 5 produces complex number values for roots x_1 and x_2, and in turn would produce values for t = exp(x) that are not positive real numbers. We therefore take x_0 as the correct solution for x = ln(t) and can describe the solution for Case 2a, including a specific exception[5] for C = 0 as:

$$t = \begin{cases} \exp(\frac{\kappa_2}{3\kappa_3}), & C = 0 \\ \exp\left(\frac{1}{3\kappa_3}\left(\kappa_2 - C - \frac{\Delta_0}{C}\right)\right), & C \neq 0 \end{cases}$$

Case 2b

- For Case 2b, calculating C requires deriving cube roots of a complex number. We first adjust the definition of C for Case 2b to be a new term C*, which extracts the imaginary number i from within the square root:

$$C^* = \sqrt[3]{\frac{\Delta_1}{2} + \frac{\sqrt{4\Delta_0^3 - \Delta_1^2}}{2} i}$$

- Although there are three complex roots for C*, we can take C* to be the "principal root," which is the root with the largest real part. To derive the cube roots, we rely on the following mathematical theorems for complex numbers[6]:

Euler's theorem: $\exp(i\theta) = \cos(\theta) + i\sin(\theta)$

De Moivre's theorem: $(\cos(\theta) + i\sin(\theta))^u = \cos(u\theta) + i\sin(u\theta)$

Polar form of a complex number: $g + hi = r\exp(i\theta)$

$$r = \sqrt{g^2 + h^2}$$

$$\theta = \arctan\left(\frac{h}{g}\right)$$

- Using the theorems above, for any complex number $g + hi$ and integer u, the u roots can be described in the following form:

$$r^{\frac{1}{u}}\left(\cos\left(\frac{\theta + 2w\pi}{u}\right) + i * \sin\left(\frac{\theta + 2w\pi}{u}\right)\right), w \in \{0,1,\dots,u-1\} \quad \text{(Equation 6)}$$

- However, the definition of θ above is a "naïve" definition which fails when the real part "g" of $g + hi$ is negative or when $g = 0$.[7]

Setting $g = \frac{\Delta_1}{2}$ and $h = \frac{\sqrt{4\Delta_0^3 - \Delta_1^2}}{2}$, and with $h > 0$ given the use of the positive square root of $\Delta_1^2 - 4\Delta_0^3$, we instead redefine θ as:

$$\theta = \begin{cases} \arctan\left(\frac{\sqrt{4\Delta_0^3 - \Delta_1^2}}{\Delta_1}\right), & \Delta_1 > 0 \\ \arctan\left(\frac{\sqrt{4\Delta_0^3 - \Delta_1^2}}{\Delta_1}\right) + \pi, & \Delta_1 < 0 \\ \frac{\pi}{2}, & \Delta_1 = 0 \end{cases}$$

- Using Equation 6 with $u = 3$, $g = \frac{\Delta_1}{2}$ and $h = \frac{\sqrt{4\Delta_0^3 - \Delta_1^2}}{2}$, and knowing θ is defined on the range 0 to π as well as the properties of cosine functions, we can determine the principle root and take C* to be:

$$C^* = \sqrt{\Delta_0} * \left(\cos\left(\frac{\theta}{3}\right) + i * \sin\left(\frac{\theta}{3}\right)\right)$$

- Feeding C* (instead of "C") along with our other values for a, b, c and d into Equation 5, we derive three real solutions for x and t = exp(x). As S(t) is a non-increasing survival function, one of these three solutions for t must match our given survival estimate $\hat{S}$ when plugged back into our function S(t), and therefore we characterize the solution for Case 2b:

$$t = \hat{t} \mid S(\hat{t}) = \hat{S}, \hat{t} \in \{\hat{t}_0, \hat{t}_1, \hat{t}_2\}$$

$$\hat{t}_0 = \exp\left[\frac{1}{3\kappa_3}\left(\kappa_2 - 2\sqrt{\Delta_0}\cos\left(\frac{\theta}{3}\right)\right)\right]$$

$$\hat{t}_1 = \exp\left[\frac{1}{3\kappa_3}\left(\kappa_2 + \sqrt{\Delta_0}\left(cos\left(\frac{\theta}{3}\right) + \sqrt{3}\,sin\left(\frac{\theta}{3}\right)\right)\right)\right]$$

$$\hat{t}_2 = \exp\left[\frac{1}{3\kappa_3}\left(\kappa_2 + \sqrt{\Delta_0}\left(cos\left(\frac{\theta}{3}\right) - \sqrt{3}\,sin\left(\frac{\theta}{3}\right)\right)\right)\right]$$

Case 3 Solution

- For Case 3, we have $\hat{S} < S(\exp(k_{max}))$ and therefore $x > k_{max}$. As indicated above, based on how Royston/Parmar spline functions are designed with linearity constraints before and after the last knots, the x^3 and x^2 terms in $v_j(x)$ from Equation 2 cancel out to 0, leaving a linear equation to solve for x. However, unlike Case 1, additional linear and constant terms are still generated from each $v_j(x)$, which when combining with the relevant γ parameter terms can be generalized to:

$$\sum_{q=2}^{n+1} \gamma_q v_{q-1}(x), \text{ x terms: } 3\sum_{q=2}^{n+1} \gamma_q (k_{q-1}^2 - \lambda_{q-1} k_{min}^2 - (1 - \lambda_{q-1}) k_{max}^2) = \tau_1$$

$$\sum_{q=2}^{n+1} \gamma_q v_{q-1}(x), \text{ constant terms: } -\sum_{q=2}^{n+1} \gamma_q (k_{q-1}^3 - \lambda_{q-1} k_{min}^3 - (1 - \lambda_{q-1}) k_{max}^3) = -\tau_0$$

- Combining these with the terms described in Equation 4 for Case 1, this then gives us the Case 3 solution as:

$$t = \exp\left(\frac{F^{-1}(\hat{S}) - \gamma_0 + \tau_0}{\gamma_1 + \tau_1}\right)$$

Results

- An analytical solution to inverse functions for all case types was derived for Royston/Parmar spline models to help facilitate direct simulation of event times, excluding covariates.

Conclusions

- Analytical solutions to inverse functions of Royston/Parmar RCS models can be derived to allow precise estimation of event times from given survival estimates, and may be useful for faster calculation of event times for DES models vs. numerical methods.
- Further research is required to update the solution to factor in the impact of covariates. While the solution may be simple to extend in the case of non-time-dependent covariates, the solution is likely more complex in the case of time-varying covariates which involve additional spline terms.

References

1. Royston P, et al. Stat Med. Aug 15 2002;21(15):2175-97. 2. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. http://www.nicedsu.org.uk. 3. Rutherford M, et al. NICE DSU Technical Support Document 21: Flexible methods for survival analysis. 2020. http://www.nicedsu.org.uk. 4. Palmer S, et al. Value Health. Feb 2023;26(2):185-192. 5. Wikipedia contributors. Cubic equation. Wikipedia, The Free Encyclopedia. September 30, 2024, 19:55 UTC. https://en.wikipedia.org/w/index.php?title=Cubic_equation&oldid=1248665266. Accessed October 17, 2024. 6. Newcastle University. Polar Form and De Moivre's Theorem. https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/core-mathematics/pure-maths/algebra/polar-form-and-de-moivre-s-theorem.html. Accessed October 17, 2024. 7. Newcastle University. Modulus and Argument. Available at: https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/core-mathematics/pure-maths/algebra/modulus-and-argument.html. Accessed October 17, 2024.

Disclosures/Acknowledgments

We would like to thank Venediktos Kapetanakis of Evidera, a business unit of PPD, part of Thermo Fisher Scientific for providing feedback on the formal write up of the solution and Michael Grossi and Shani Berger of Evidera, a business unit of PPD, part of Thermo Fisher Scientific for their editorial and graphic design contributions.

GB and JM are employees of Evidera. The views expressed in this study are those of the authors and not necessarily those of Evidera. The study was not funded by any organization. The authors do not have any conflicts of interest.

Presented at the ISPOR Europe Conference • 17–20 November 2024 • Barcelona, Spain

Funding provided by Evidera Ltd., a business unit of PPD, part of Thermo Fisher Scientific