



Andre Verhoek<sup>\*,1</sup> , Mario JNM Ouwens<sup>2</sup>, Bart Heeg<sup>3</sup> , Frank GA Jansman<sup>4,5</sup>   
& Maarten J Postma<sup>1,6,7,8</sup> 

<sup>1</sup>Unit of Global Health, Department of Health Sciences, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

<sup>2</sup>Real World Science & Analytics, AstraZeneca Global, Västergötland, Sweden

<sup>3</sup>Cytel, Rotterdam, The Netherlands

<sup>4</sup>Department of Clinical Pharmacy, Deventer Teaching Hospital, Deventer, The Netherlands

<sup>5</sup>Unit of Pharmacotherapy, Epidemiology & -Economics, Groningen Research Institute of Pharmacy, University of Groningen, Groningen, The Netherlands

<sup>6</sup>Department of Economics, Econometrics & Finance, Faculty of Economics & Business, University of Groningen, The Netherlands

<sup>7</sup>Center of Excellence for Pharmaceutical Care Innovation, Universitas Padjadjaran, Bandung, Indonesia

<sup>8</sup>Division of Pharmacology & Therapy, Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia

\*Author for correspondence: [d.verhoek@umcg.nl](mailto:d.verhoek@umcg.nl)

**Aim:** Fractional Polynomial (FP) models are widely used in survival analysis for health technology assessment and network meta-analysis (NMA). However, current implementations rely on a fixed set of pre-specified powers, which may constrain model flexibility, limit predictive performance and increase computational cost in Bayesian settings. This study introduces and evaluates a Bayesian FP modeling approach in which the powers are estimated as continuous parameters rather than fixed, aiming to simplify model selection and improve fit. **Materials & methods:** Second-order Bayesian FP models were implemented in STAN, allowing the time transformation powers ( $p_1$ ,  $p_2$ ) to be estimated from the data. Model performance was evaluated across three oncology NMA datasets; in advanced non-small-cell lung cancer, metastatic prostate cancer and early breast cancer. The performance was assessed using visual fit, leave-one-out-information-criteria, root mean square error, incremental survival estimates and computational efficiency. Validation steps included posterior predictive checks, sensitivity analyses and long-term extrapolation. **Results:** Across all datasets, variable power models consistently achieved better statistical fit (lower leave-one-out-information-criteria and root mean square error) than fixed power models. Incremental survival estimates were also more stable and clinically plausible, particularly in datasets with complex hazard dynamics. While variable models required slightly more time per run, the approach greatly reduced the number of required model configurations, leading to lower overall computational burden. **Conclusion:** Bayesian FP models with variable powers not only improve model fit and simplify model selection but also reduce structural uncertainty by replacing exhaustive grid searches with a unified, data-driven estimation of transformation powers, while retaining interpretability and computational efficiency. By producing robust, well-calibrated survival projections and streamlining model selection, this approach strengthens survival analysis for health technology assessment and supports more reliable decision-making in comparative effectiveness research.

First draft submitted: 5 August 2025; Accepted for publication: 7 November 2025; Published online: 9 December 2025

**Keywords:** Bayesian network meta-analysis • computational efficiency • fractional polynomial • health technology assessment • structural uncertainty • survival extrapolation

In survival analysis, accurately extrapolating survival functions beyond the observed data is crucial for making informed decisions in clinical and health policy settings. Traditional parametric distributions, while useful, often lack the flexibility needed to capture complex, nonlinear, time-varying hazards in survival data. This has led to the development of more flexible modeling techniques, such as fractional polynomials (FPs).

First introduced by Royston and Parmar [1] in 1994, FPs offer a flexible alternative to traditional polynomial regression, allowing for the modeling of nonlinear relationships in survival analysis. Later, in 2008, Royston and Sauerbrei [2] explained that their canonical power set  $[-2, -1, -0.5, 0, 0.5, 1, 2, 3]$  was chosen “for practical reasons (interpretability, computational stability, reproducibility)” and noted that “in principle, the set could be enlarged or modified, but . . . the proposed set gives satisfactory results in practice”. However, this requires running all combinations of the powers, which can be time-consuming in a Bayesian framework.

In 2011, Jansen [3] extended the application of FPs to network meta-analysis (NMA) within a Bayesian framework, significantly broadening the scope of these models in comparative effectiveness research. FPs are also referenced in the National Institute for Health and Care Excellence Technical Support Document 21 (TSD 21) [4] as one of several parametric approaches for extrapolating survival data. Although not explored in depth, their inclusion highlights the growing recognition of flexible modeling techniques in health economic evaluation. Since then, FPs have gained widespread acceptance in health technology assessment (HTA) submissions, becoming a standard tool in the evaluation of medical interventions [5–9].

Despite their widespread adoption, FP models have remained anchored to that original fixed power set – a pragmatic rather than theoretical necessity [1,2,10]. Advances in computational power and the introduction of STAN, which outperforms BUGS and JAGS for Hamiltonian Monte Carlo sampling, may remove the original constraints on power selection [11]. This evolution prompts the question: might FP models perform better if powers were treated not as fixed points but as continuous estimable parameters? Allowing powers as estimable parameters would enable a finer-grained transformation of time, potentially improving model fit, producing more clinically plausible extrapolations and yielding more realistic estimates of incremental survival between treatments.

Heeg *et al.* [12] already mentioned the use of variable powers in FPs as part of a package of novel methods, but did not elaborate on the methodology. To fill this gap, we implement and evaluate variable-power FP models in RStan – publishing, for the first time, complete STAN code for FP survival models in an NMA setting – and compare them against the traditional fixed-power approach across three cancer NMA datasets. In addition to statistical fit, we assess the benefits of variable powers for clinically plausible extrapolations, incremental survival estimates, computational efficiency and robust posterior validation. Through this work, we aim to simplify and extend the FP framework, harnessing modern Bayesian tools to deliver more precise and reliable survival predictions in comparative effectiveness research.

## Methodology

### Setting the bayesian framework

This study applies a Bayesian framework to model time-to-event data using FPs within both survival analysis and NMA. FPs offer a flexible approach to modeling nonlinear time-varying hazards, which is crucial for accurate extrapolation in survival analysis and comparative effectiveness research.

#### General FP formulation

The core FP model for the hazard function  $h(t)$  in a survival setting is defined, per Jansen [3], as

$$\log(h(t)) = \beta_0 + \sum_{m=1}^M \beta_m t^{p_m} \quad (\text{Equation 1})$$

where:

- $h(t)$  is the hazard at time  $t$ ,
- $\beta_0$  is the baseline log-hazard,
- $M$  is the order of the FP
- $\beta_m$  are the coefficients for the polynomial terms,
- $t^{p_m}$  represents time raised to power  $p_m$ .

In most applications, the powers  $p_m$  are selected from a predefined set (commonly  $[-2, -1, -0.5, 0, 0.5, 1, 2, 3]$ ) chosen to balance flexibility and computational feasibility. However, this choice is not fundamental to the method.

In a Bayesian setting, all model parameters, including  $\beta_0$ ,  $\beta_m$  and potentially the powers  $p_m$ , are treated as random variables with prior distributions. This framework allows the analyst to either fix the powers (as traditionally done) or estimate them directly from the data by assigning priors, thus enabling a variable power formulation. The

posterior distributions of all parameters are then derived via Bayesian inference using Markov Chain Monte Carlo (MCMC) methods. While the formulation in Equation 1 allows for any FP order (i.e.,  $M \geq 1$ ), in practice, only first- and second-order FPs are typically used [5–9].

### Extending FPs to NMA

In an NMA context, where multiple treatments are compared across different studies, the FP model is extended by incorporating study-specific and treatment-specific random effects [3]. The model is formulated as:

$$\log(h_{ijk}(t)) = \beta_{0,j,k} + \sum_{m=1}^M \beta_{m,j,k} t^{p_m} \quad (\text{Equation 2})$$

where  $h_{ijk}(t)$  represents the hazard function for the  $i$ th individual in study  $j$  receiving treatment  $k$ ,  $\beta_{0,j,k}$  is the study-specific and treatment-specific baseline log-hazard,  $\beta_{m,j,k}$  are the coefficients for the FP terms that vary by study and treatment,  $t^{p_m}$  are the FP terms as before, with  $p_m$  either fixed or treated as variables.

To ensure flexibility, the model accommodates a hierarchical structure where the coefficients  $\beta_{0,j,k}$  and  $\beta_{m,j,k}$  are modeled using normal distributions with hyperparameters reflecting prior uncertainty:

$$\beta_{0,j,k} \sim \mathcal{N}(\mu_{0,j,k}, \tau_{0,j,k}) \text{ and } \beta_{m,j,k} \sim \mathcal{N}(\mu_{m,j,k}, \tau_{m,j,k}) \text{ for } m \in \{1, \dots, M\} \quad (\text{Equation 3})$$

where  $M$  is the order of the FP, and  $\mu_{0,j,k}$  and  $\mu_{m,j,k}$  are the mean effects for study  $j$  and treatment  $k$ , and  $\tau_{0,j,k}$  and  $\tau_{m,j,k}$  are precision parameters (inverse covariances). These hyperparameters themselves can have prior distributions, reflecting the uncertainty in the variation between studies and treatments.

In the current specification, treatment and study effects are defined for each parameter ( $\beta_0, \beta_1, \beta_2$ ). However, the framework can be adapted to restrict treatment effects to only the scale parameter ( $\beta_0$ ) or to a subset of shape parameters as desired. Previous research has shown that placing treatment effects on all FP parameters may risk overfitting, particularly in small or sparse networks (Freigofaite *et al.*) [13]. For the purposes of this paper, we retained the full specification across all parameters in order to focus the comparison squarely on the difference between fixed versus variable powers. Future applications could tailor the placement of treatment effects depending on data availability, model stability and clinical plausibility.

Finally, to ensure long-term survival estimates reflect background mortality, we applied a post-fit adjustment using national life tables [14]. Population survival  $S_{pop}(t)$  were interpolated to monthly intervals used the approach by Drummond *et al.* [15] and combined with the model-based survival as described above.

### Priors & posterior inference

To complete the Bayesian model, we specify prior distributions for the parameters:

- For  $\beta_{0,j,k}$  and  $\beta_{m,j,k}$ , we use weakly informative normal priors, such as  $\mathcal{N}(0, 2)$  to allow the data to drive the inference while still regularizing the model. Note that the  $\beta_{m,j,k}$  parameters vary over time, and in this model, time is expressed in monthly units.

The posterior distributions of the parameters are estimated using MCMC methods. In this study, all models are implemented in STAN using Hamiltonian Monte Carlo, which provides efficient sampling and robust convergence diagnostics.

### Variable powers: method & FP formulation

Building upon the traditional FP model, which uses a fixed set of powers, we introduce the concept of variable powers for enhanced flexibility in modeling nonlinear time-varying hazards. By allowing powers to vary rather than remain fixed, we increase the model's ability to adapt to the complexity of survival data, especially in cases with intricate hazard trends. This extension naturally follows the basic FP formulation and represents more flexibility by providing a wider range of polynomial transformations, potentially leading to better model fit and interpretability.

### Extending fixed powers to variable powers

In traditional FPs, the powers  $p_m$  are selected from a predefined set (e.g.,  $\in [-2, -1, -0.5, 0, 0.5, 1, 2, 3]$ ) designed to balance flexibility with computational feasibility [16]. However, advances in computational power and Bayesian sampling techniques now allow us to relax this constraint. In the variable power approach, the powers  $p_m$  are treated

as random variables, enabling them to be directly estimated from the data rather than constrained to fixed values. This shift requires introducing priors for  $p_m$ , which allows the Bayesian framework to estimate them within the MCMC process, adapting to the data's structure dynamically.

#### *Choice of prior distributions for variable powers*

Given the unknown nature of the powers  $p_m$ , we select weakly informative priors, specifically  $N(0, 2)$ , as the initial prior, offering flexibility and interpretability [17]. By assuming a normal prior, we can center the distribution around plausible values, capturing the expected range of  $p_m$  while allowing the data to inform their exact values. However, should posterior diagnostics (such as plotting the joint posterior distributions) suggest deviation from normality, alternative priors, such as uniform (if no prior knowledge is assumed) or student's  $t$ -distribution (to accommodate heavier tails), can be considered.

#### *New formulations with variable powers*

The idea of estimating FP powers as continuous rather than fixed parameters has previously been explored. Heeg *et al.* demonstrated an implementation of variable powers in an NMA context, but did not provide a formal mathematical formulation [12]. Moreover, their implementation followed a different specification from the one we propose here. In this work, we present the first complete Bayesian specification and Stan code of variable-power FP survival models in an NMA setting, thereby extending prior work into a stable and reproducible modeling framework.

The revised FP model with variable powers adjusts the hazard function to

$$\log(h_{ijk}(t)) = \beta_{0,j,k} + \sum_{m=1}^M \beta_{m,j,k} t^{p_m} \text{ where } p_m \geq p_{m-1} \forall m \in (M \wedge m \neq 1) \quad (\text{Equation 4})$$

where  $h_{ijk}(t)$  is the hazard for individual  $i$  in study  $j$  and treatment  $k$  at time  $t$ . Here,  $M$  denotes the order of the fractional polynomial, and  $m$  indexes the FP terms. Unlike traditional FPs, where  $p_m$  is selected from a discrete grid, we allow  $p_m$  to be estimated directly from the data as random variables with prior distributions. This increases modeling flexibility by enabling finger-grained time transformations, at the cost of introducing additional parameters.

A difficulty arises from the behavior of FP transformations near zero. Powers close to zero can effectively 'flip' behavior (e.g.,  $t^{0.1}$  vs  $t^{-0.1}$ , see [Supplementary Figure 1](#) in the Appendix as example), complicating the optimization of the powers as random, continuous variables. To address this, we structured the estimation into three distinct model configurations, each covering one region of the solution space:

- $p_1 < 0, p_2 < 0$ : both powers constrained to negative values;
- $p_1 < 0, p_2 > 0$ : one power constrained negative, one positive;
- $p_1 > 0, p_2 > 0$ : both powers constrained to positive values.

This separation avoids unstable switching behavior around zero, while allowing flexibility across clinical relevant shapes of survival curves.

We also imposed an ordering constraint, such that  $p_m \geq p_{m-1}$  for all  $m \neq 1$ . Without this, solutions such as ( $p_1 = 1$  and  $p_2 = 2$ ) and ( $p_1 = 2$  and  $p_2 = 1$ ) are mathematically equivalent but would be treated as distinct in posterior sampling, leading to multiple symmetric optima. Such redundancy complicates estimation, risks label-switching behavior and may generate misleading multimodal posteriors. Enforcing an order ensures uniqueness of solutions and stabilizes inference.

Theoretically, powers in FP models can be any real number ( $-\infty < p_m < \infty$ ). In practice, however, extreme values are not meaningful because survival functions are defined through a double exponential form

$$s(t) = \exp(-\exp(\log(h(t))))$$

large absolute powers drive hazards to extreme values, resulting in degenerate survival probabilities (collapsing toward 0 or 1). While the Bayesian estimation process, guided by priors and data, naturally avoids implausible regions, optimization is also stabilized by restricting powers to a reasonable range (e.g.,  $-3 \leq p_m \leq 3$ ).

## Introducing the data

This study uses data from three published NMAs, each focused on evaluating the comparative effectiveness of multiple treatments in advanced cancer settings. The datasets were selected for their clinical relevance, richness in survival outcomes and availability of long-term follow-up data suitable for validating extrapolative modeling. The Kaplan–Meier (KM) of each relevant treatment arm in the studies of the three networks (depicted in Figures 1–3) are digitized using Engauge Digitizer [18] and transformed into pseudo-individual patient level data using the algorithm by Guyot *et al.* [19]. To maintain clarity and focus, a subset of treatment arms was selected for each network based on clinical relevance and completeness of follow-up. The three NMAs, and associated validation studies, are summarized below.

### Network 1: advanced non-small-cell lung cancer

The first dataset is based on a systematic literature review and NMA comparing nivolumab plus ipilimumab to other first-line therapies in advanced non-small-cell lung cancer (aNSCLC) (O’Byrne *et al.*) [20]. From this NMA, we selected and digitized overall survival (OS) curves from the following trials: IMpower150, KEYNOTE-189 and CheckMate-227 studies involving combinations of chemotherapy, immune checkpoint inhibitors and anti-angiogenics (Figure 1). We focused on arms that included Nivo + Ipi (nivolumab + ipililumab), ABCP (atezolizumab + bevacizumab + chemo + platinum) and pembrolizumab-based regimens. From the underlying studies of the NMA, the IMPower150 trial (Socinski *et al.*) and KEYNOTE-189 (Garassino *et al.*) provide extended OS follow-up for two of the treatment arms and are used to validate long-term extrapolations [21,22].

### Network 2: metastatic castration-sensitive prostate cancer

The second dataset comes from an NMA by Wang *et al.* [23], comparing systemic therapies for metastatic castration-sensitive prostate cancer. For feasibility and consistency in follow-up, we selected the ARCHES and TITAN trials. These provide radiographic progression-free survival outcomes for treatment arms including ADT alone, ADT + docetaxel and ADT + enzalutamide (Figure 2). From the underlying studies of the NMA, the ARCHES trial (Armstrong *et al.*) [24], which reported updated follow-up by metastatic pattern, is used to validate long-term predictions in this network.

### Network 3: HER2-positive early breast cancer

The third dataset is based on an NMA protocol published by Wilson *et al.* [25] concerning HER2-positive early breast cancer (eBC). We selected the BrighTNess trial, CALGB 40603 (parts a and b), and KEYNOTE-522, targeting arms with trastuzumab combined with varying chemotherapy regimens, including paclitaxel + carboplatin + veliparib (P + C + V) and regimens containing bevacizumab. Event-free survival was selected as the outcome due to its clinical relevance in early-stage disease (Figure 3). Among these, the KEYNOTE-522 trial (Koonstra *et al.*) [26] provides extended follow-up for pembrolizumab containing regimens to assess extrapolation performance.

Across all networks, the long-term follow-up datasets were not used in model fitting, but solely to assess the predictive performance of each modeling approach beyond the original data horizon. This distinction enables a realistic evaluation of how well each model extrapolates, a critical requirement in survival modeling for cost-effectiveness and policy analysis.

These three indications were deliberately selected to test the FP approach under diverse clinical dynamics. aNSCLC is a highly aggressive malignancy with steep early declines in survival and potential immunotherapy tail effects, requiring flexible modeling of nonproportional hazards. In contrast, metastatic castration-sensitive prostate cancer typically follows a more gradual progression, with delayed treatment effects and extended disease control, challenging models to capture subtle nonlinear hazard patterns. HER2-positive eBC, meanwhile, presents a curative intent setting where relapse plateaus over time, making it essential to model long-term remission accurately. Together, these networks span a broad spectrum of survival trajectories and extrapolation demands, providing a robust test bed for evaluating the adaptability of variable-power FP models in both aggressive and indolent disease contexts.

## Comparison criteria

In this section, we detail how we compare the performance of the Bayesian FP models described above. Each comparison aims to provide insights into model accuracy, computational efficiency and the suitability of each

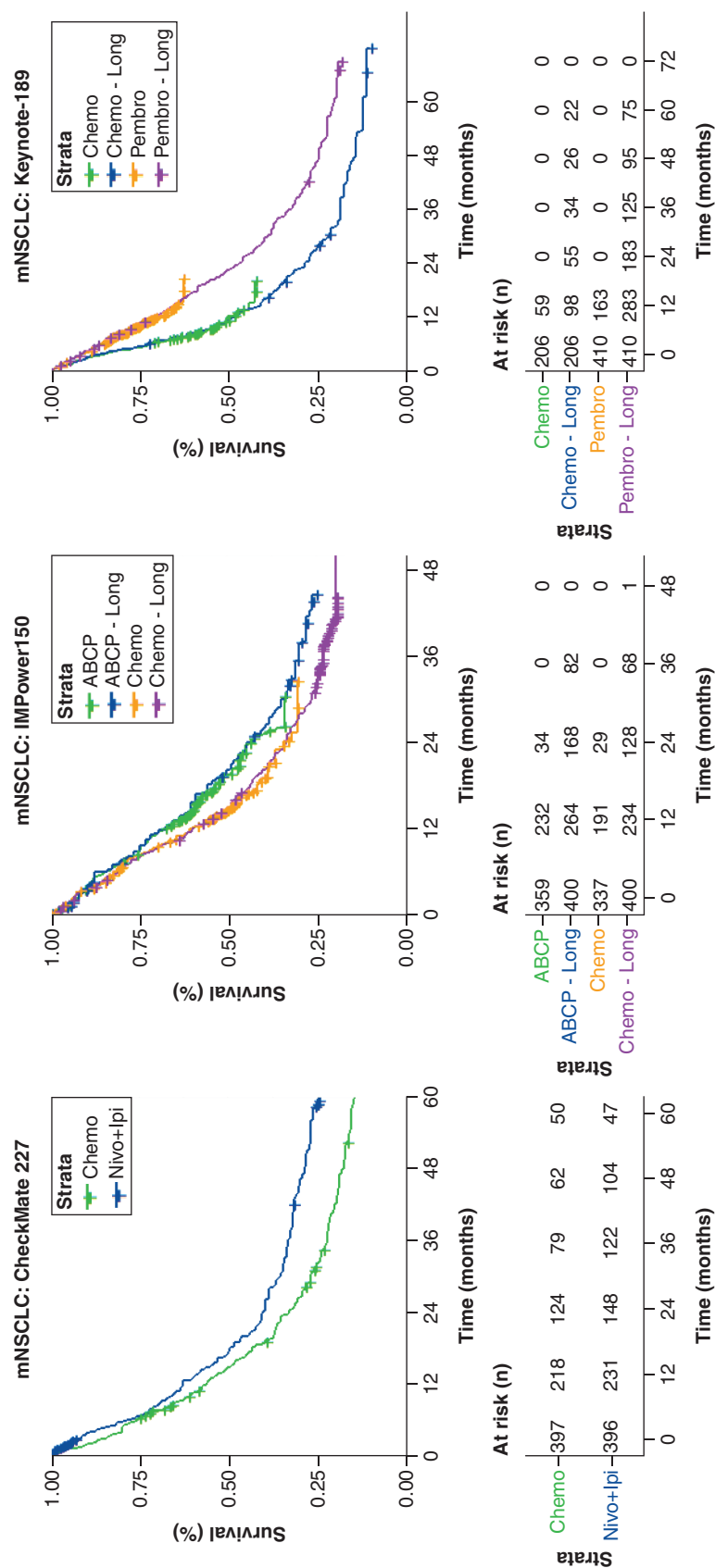


Figure 1. Observed Kaplan–Meier estimated for each study arm in the advanced non-small-cell lung cancer network, shown per study. ABCP: Atezolizumab, bevacizumab, carboplatin and paclitaxel.

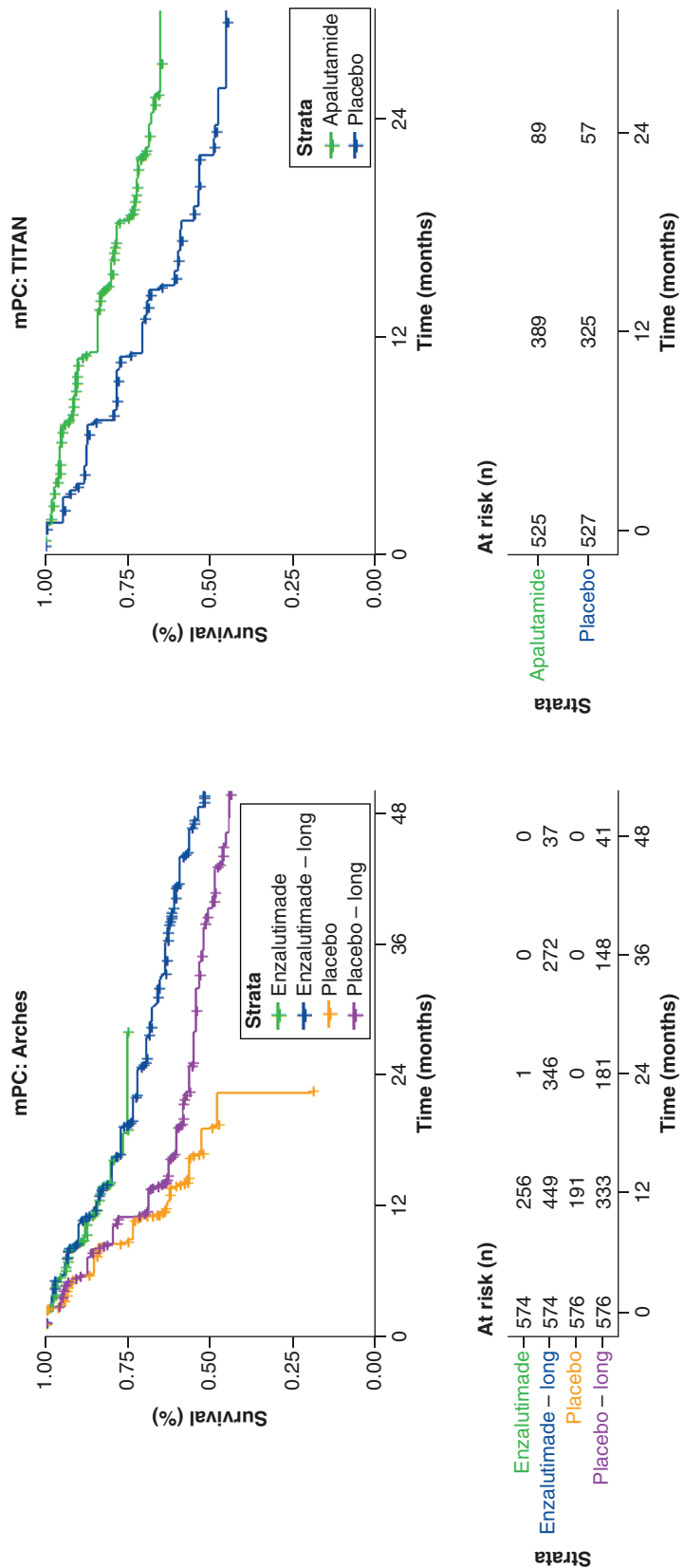


Figure 2. Observed Kaplan-Meier estimated for each study arm in the metastatic prostate cancer network, shown per study. mPC: Metastatic prostate cancer.

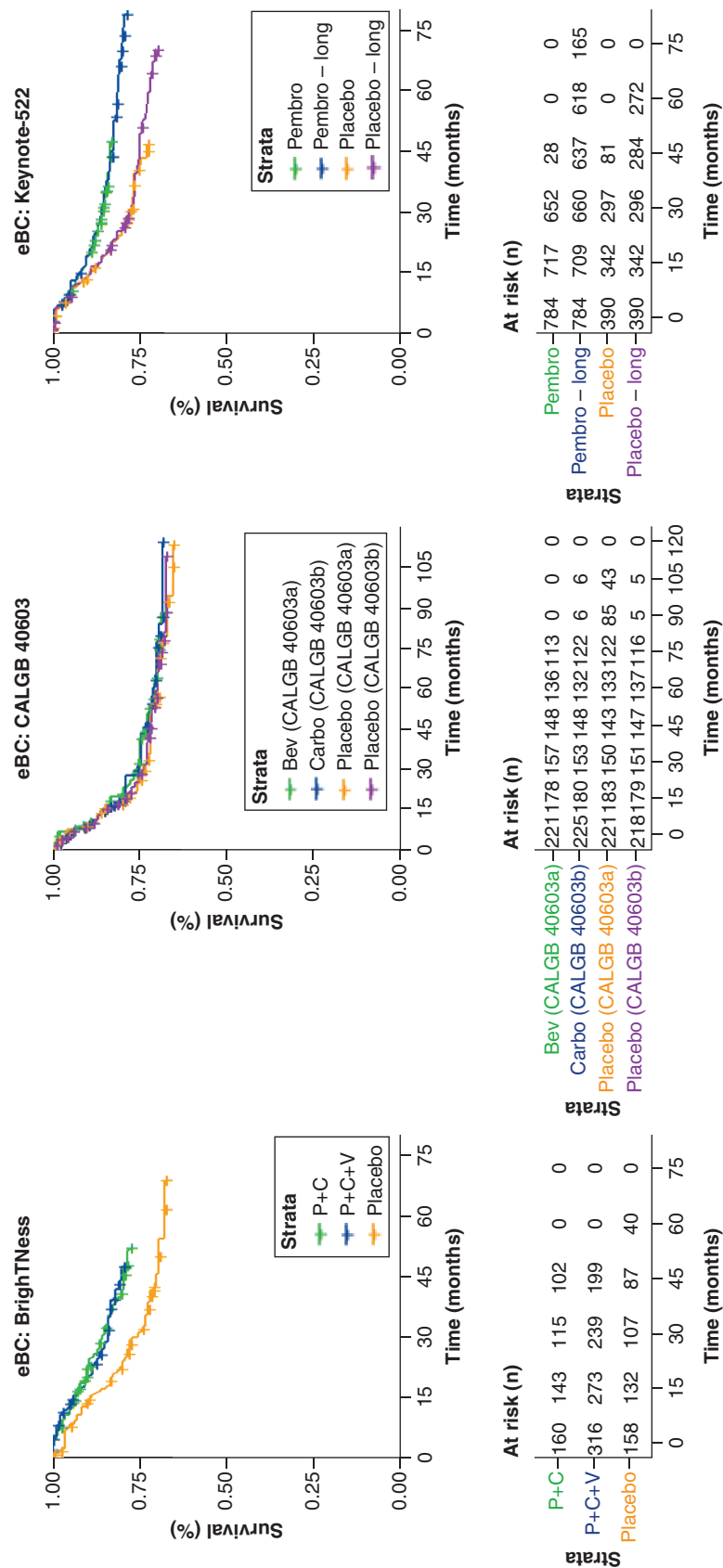


Figure 3. Observed Kaplan–Meier estimated for each study arm in the early breast cancer network, shown per study. eBC: Early breast cancer.

**Table 1.** The statistically top five models, for each dataset, based on leave-one-out-information-criteria.

| Rank | eBC                         |       |       |         | mPC                         |       |       |         | mNSCLC                      |       |      |           |
|------|-----------------------------|-------|-------|---------|-----------------------------|-------|-------|---------|-----------------------------|-------|------|-----------|
|      | Model                       | p1    | p2    | LOOIC   | Model                       | p1    | p2    | LOOIC   | Model                       | p1    | p2   | LOOIC     |
| 1    | Variable (-/+) <sup>†</sup> | -0.74 | 0.33  | 7452.00 | Variable (-,-) <sup>†</sup> | -1.63 | -0.75 | 6083.20 | Variable (-/+) <sup>†</sup> | -0.79 | 0.24 | 10,373.60 |
| 2    | Fixed                       | -0.5  | 0.5   | 7481.40 | Variable (-/+) <sup>†</sup> | -1.08 | 0.12  | 6088.30 | Fixed                       | -1    | 1    | 10,388.10 |
| 3    | Variable (-,-) <sup>†</sup> | -1.2  | -0.26 | 7482.70 | Fixed                       | -2    | 0.5   | 6094.40 | Fixed                       | -1    | 0.5  | 10,389.70 |
| 4    | Fixed                       | -1    | 1     | 7490.70 | Fixed                       | -2    | -0.5  | 6095.30 | Fixed                       | -0.5  | 1    | 10,391.40 |
| 5    | Fixed                       | -1    | 0.5   | 7490.90 | Fixed                       | -2    | -1    | 6096.40 | Fixed                       | -0.5  | 0.5  | 10,393.30 |

<sup>†</sup>The type of variable model is depicted between brackets.

eBC: Early breast cancer; LOOIC: Leave-one-out-information criterion; mPC: Metastatic prostate cancer.

approach for handling complex survival data in NMA settings. Therefore, we set the following criteria, as listed below.

### Comparison criteria

The primary criteria for comparison are visual fit, statistical fit, computational efficiency and long-term predictive validation:

- **Statistical fit:** We evaluate model fit using the leave-one-out-information criterion (LOOIC), which is particularly suited for Bayesian models [27]. This metric accounts for model complexity and provide an estimate of predictive accuracy. Lower LOOIC values indicate superior predictive accuracy, taking model complexity into account. Additionally, we calculate the root mean square error (RMSE) to quantify the model's ability to predict survival outcomes accurately, comparing predicted values with observed data.
- **Visual fit:** We assess the visual fit by overlaying model-predicted survival curves on the observed KM estimates for each treatment arm and study. For each fitted model, we generate posterior mean survival curves and compare these visually to the empirical KM data over the available follow-up period. A close alignment between the predicted survival curve and the observed KM points is considered indicative of good model fit.
- **Computational efficiency:** To assess the efficiency of each implementation, we measure the time required for convergence. All models are run with three chains and 5000 iterations, of which 2000 are to be used as burn-in.
- **Long-term predictive validation:** To judge the extrapolation quality, we use later-published follow-up data (2–3 years beyond the digitized period) for each network, where available and compare these on visual fit and statistical fit (RMSE).
- **Uncertainty:** We examine predictive uncertainty both qualitatively and quantitatively. While variable models introduce additional parameters (which can increase posterior uncertainty), they often reduce structural uncertainty by simplifying model selection (e.g., reducing running all power set combinations to three variable formulations). Additionally, we inspect the posterior distributions of key parameters, highlighting how uncertainty is distributed and how it might propagate into downstream analyses.

By combining these qualitative and quantitative assessments, we rigorously evaluate each model's ability to forecast long-term survival, critical for robust health-economic modeling and policy decision-making. Detailed diagnostics are provided in the Appendix.

By applying these comparison methods and validation steps, we systematically evaluate each model configuration, providing insights into which combination offers the best balance of predictive accuracy, efficiency and interpretability for NMA survival analysis. This comparative analysis is essential for identifying the most effective FP model configuration for capturing nonlinear survival trends and supporting comparative effectiveness research in oncology and other indications.

Through these methods, the study aims to present a comprehensive evaluation framework, providing evidence for the strengths and limitations of each modeling approach in the context of complex survival data.

## Results

### Statistical fit

Table 1 summarizes the top five models for each of the three datasets – eBC, metastatic prostate cancer (mPC) and mNSCLC – ranked by their predictive performance as measured by LOOIC.

Notably, the best performing model in every dataset is a variable power model, which consistently achieves a substantially lower LOOIC compared with its fixed counterparts. For example, in the eBC dataset, the top model – a variable model with a negative-positive configuration ( $p_1 = -0.74$ ,  $p_2 = 0.33$ ) – yields a LOOIC of 7452.0. Similarly, for the mPC dataset, the optimal model is a variable model with both negative powers ( $p_1 = -1.63$ ,  $p_2 = -0.75$ ; LOOIC = 6083.2), and for the aNSCLC dataset, the leading model is again a variable model with a Negative-Positive configuration ( $p_1 = -0.79$ ,  $p_2 = 0.24$ ; LOOIC = 10,373.6).

Importantly, the optimized  $p_1$  and  $p_2$  values derived from the variable models generally fall between the best estimates available from the fixed grid. This observation supports our hypothesis that the variable approach, allowing continuous tuning of the powers, effectively fine-tunes the time transformation, capturing the underlying survival dynamics more precisely than the discrete, fixed set. The RMSE values, which will be finalized in subsequent analyses, offer an additional layer of validation for these models. Further details and comprehensive comparisons across all evaluated models are provided in the Appendix.

Table 2 presents the RMSE comparisons between variable and fixed effect models across the three datasets. The comparisons are shown separately for the short-term data (on which the models were originally fitted) and the long-term data (for external validation).

Consistent with the LOOIC results reported in Table 1, variable power models generally exhibit improved predictive accuracy, as reflected by lower RMSE values. In the mNSCLC dataset, the short-term RMSE is substantially reduced under the variable model (0.0187 vs 0.1579 for fixed), indicating a closer alignment with the observed survival curves during the modeled period. Similarly, in the mPC dataset, variable models consistently outperform fixed models on both short- and long-term RMSE, although the magnitude of improvement varies by treatment.

In the eBC dataset, however, RMSE differences are negligible. Both variable and fixed models show similarly low error for short-term predictions (0.0083 vs 0.0080) and identical RMSE (0.0052) for long-term follow-up. This suggests that, for this relatively stable survival pattern, the added flexibility of the variable model does not substantially improve predictive accuracy.

These findings reinforce the model selection patterns observed with LOOIC and demonstrate that variable power models consistently match or outperform fixed approaches across diverse clinical contexts.

### Incremental survival

Table 3 reports the total incremental survival across all time points for each treatment arm, comparing the variable and fixed models within each dataset. Values are presented alongside their associated standard deviations, reflecting uncertainty in the cumulative estimates.

In the aNSCLC dataset, variable models produced more plausible and stable estimates of incremental survival compared with fixed models. For example, under the variable model, the ABCP treatment in CheckMate 227 achieved an incremental survival of 1.78 (standard deviation (SD) 0.72) compared with 17.95 (SD 1.85) under the fixed model, suggesting substantial overestimation in the fixed configuration. Similarly, for Nivo + Ipi, the variable model yielded an incremental survival of 44.05 (SD 0.67) versus 38.21 (SD 1.38) under the fixed model, indicating closer but still meaningful differences.

In the mPC dataset, the variable models again showed better control over extreme predictions. For example, Apalutamide in the Arches study achieved an incremental survival of 70.25 (SD 2.66) with the variable model, whereas the fixed model estimated a substantially higher 195.34 (SD 4.74). Conversely, smaller differences were observed for treatments such as enzalutamide, with more moderate discrepancies between variable and fixed models.

In the eBC dataset, differences between variable and fixed model incremental survival estimates were more nuanced. While many treatments demonstrated similar results across models, certain arms (such as pembrolizumab in the Keynote-522 study) still showed large differences, with the variable model producing higher cumulative survival (100.8, SD 1.79) compared with the fixed model (77.39, SD 1.30).

### Visual fit

For each dataset, model-predicted survival curves were visually compared with the observed KM curves across all relevant treatment arms, see Figures 4–6, which show the best fit of both the fixed and variable model. The variable and fixed FP models produced nearly identical fits over the observed follow-up period but diverged notably in their extrapolated tails beyond the data horizon.

In the aNSCLC network, the extrapolations show clear differences across models, see Figure 4. In the KEYNOTE-189 study, for instance, the fixed model visibly overestimates long-term survival for pembrolizumab, diverging from

Table 2. Root mean square error comparisons between variable and fixed effect models across three datasets, for both the short-term (on which the models are fitted) and long-term data.

| Treatment     | Study         | mNSCLC  |                 |                 |                               | mPC                 |        |         |                 | eBC             |                               |               |              |         |                 |                 |                               |
|---------------|---------------|---------|-----------------|-----------------|-------------------------------|---------------------|--------|---------|-----------------|-----------------|-------------------------------|---------------|--------------|---------|-----------------|-----------------|-------------------------------|
|               |               | KM type | Variable RMSE   | Fixed RMSE      | Difference (fixed – variable) | Treatment           | Study  | KM type | Variable RMSE   | Fixed RMSE      | Difference (fixed – variable) | Treatment     | Study        | KM type | Variable RMSE   | Fixed RMSE      | Difference (fixed – variable) |
| Chemo         | CheckMate 227 | Short   | 0.0108 (0.0091) | 0.1119 (0.0651) | 0.1011                        | Enzalutamide        | Arches | Short   | 0.0272 (0.0351) | 0.0310 (0.0379) | 0.0037                        | P + C         | BrightNESS   | Short   | 0.0040 (0.0035) | 0.0037 (0.0033) | -0.0003                       |
| Nivo + Ipi    | CheckMate 227 | Short   | 0.0116 (0.0076) | 0.1408 (0.0595) | 0.1291                        | Placebo             | Arches | Short   | 0.0540 (0.0609) | 0.0521 (0.0608) | -0.0019                       | P + C + V     | BrightNESS   | Short   | 0.0090 (0.0069) | 0.0084 (0.0067) | -0.0006                       |
| ABCP          | IMPower150    | Short   | 0.0179 (0.0151) | 0.2020 (0.0585) | 0.1842                        | Apalutamide         | TITAN  | Short   | 0.0345 (0.0382) | 0.0291 (0.0352) | -0.0054                       | Placebo       | BrightNESS   | Short   | 0.0135 (0.0075) | 0.0100 (0.0062) | -0.0034                       |
| Chemo         | IMPower150    | Short   | 0.0248 (0.0255) | 0.0607 (0.0636) | 0.0359                        | Placebo             | TITAN  | Short   | 0.0987 (0.0706) | 0.1168 (0.0948) | 0.0181                        | Bev           | CALGB 40603a | Short   | 0.0076 (0.0072) | 0.0087 (0.0064) | 0.0011                        |
| Chemo         | Keynote-189   | Short   | 0.0269 (0.0271) | 0.1838 (0.0517) | 0.1569                        | Enzalutamide – Long | Arches | Long    | 0.0613 (0.0511) | 0.0903 (0.0819) | 0.029                         | Placebo       | CALGB 40603a | Short   | 0.0099 (0.0089) | 0.0106 (0.0071) | 0.0007                        |
| Pembro        | Keynote-189   | Short   | 0.0202 (0.0205) | 0.2483 (0.0346) | 0.2281                        | Placebo – Long      | Arches | Long    | 0.1177 (0.0901) | 0.1120 (0.0836) | -0.0057                       | Carbo         | CALGB 40603b | Short   | 0.0058 (0.0053) | 0.0067 (0.0056) | 0.0009                        |
| ABCP – Chemo  | IMPower150    | Long    | 0.0484 (0.0402) | 0.1957 (0.0517) | 0.1473                        |                     |        |         |                 |                 |                               | Placebo       | CALGB 40603b | Short   | 0.0069 (0.0070) | 0.0065 (0.0048) | -0.0005                       |
| Chemo – Long  | IMPower150    | Long    | 0.0404 (0.0369) | 0.0390 (0.0543) | -0.0014                       |                     |        |         |                 |                 |                               | Pembro        | Keynote-522  | Short   | 0.0044 (0.0028) | 0.0034 (0.0023) | -0.001                        |
| Chemo – Long  | Keynote-189   | Long    | 0.0212 (0.0120) | 0.0662 (0.0743) | 0.045                         |                     |        |         |                 |                 |                               | Placebo       | Keynote-522  | Short   | 0.0081 (0.0053) | 0.0073 (0.0048) | -0.0008                       |
| Pembro – Long | Keynote-189   | Long    | 0.0292 (0.0172) | 0.1136 (0.0912) | 0.0844                        |                     |        |         |                 |                 |                               | Pembro – Long | Keynote-522  | Long    | 0.0032 (0.0025) | 0.0049 (0.0044) | 0.0018                        |
| Average       |               | Short   | 0.0187          | 0.1579          | 0.1392                        | Average             |        | Short   | 0.0536          | 0.0573          | 0.0037                        | Average       | Keynote-522  | Long    | 0.0183 (0.0137) | 0.0136 (0.0089) | -0.0047                       |
|               |               | Long    | 0.0303          | 0.0729          | 0.0427                        |                     |        | Long    | 0.0895          | 0.1012          | 0.0117                        |               |              | Long    | 0.0052          | 0.0052          | 0.0000                        |

RMSE values are reported with standard deviations in parentheses.  
BCP: Atezolizumab, bevacizumab, carboplatin and paclitaxel; eBC: Early breast cancer; KM: Kaplan–Meier; mPC: Metastatic prostate cancer; RMSE: Root mean square error.

RMSE values are reported with standard deviations in parentheses.

ABCP: Atezolizumab, bevacizumab, carboplatin and paclitaxel; eBC: Early breast cancer; KM: Kaplan–Meier; mPC: Metastatic prostate cancer; RMSE: Root mean square error.

Table 3. Incremental survival in months for each dataset, shown per study and treatment using the placebo or chemo arm as benchmark.

| Study         | mNSCLC     |               |               |        | mPC          |               |               |              | eBC       |               |              |       |
|---------------|------------|---------------|---------------|--------|--------------|---------------|---------------|--------------|-----------|---------------|--------------|-------|
|               | Treatment  | Variable (SD) | Fixed (SD)    | Study  | Treatment    | Variable (SD) | Fixed (SD)    | Study        | Treatment | Variable (SD) | Fixed (SD)   | Study |
| CheckMate 227 | Chemo      | 0 (0)         | 0 (0)         |        |              |               |               | BrightTNess  | Placebo   | 0 (0)         | 0 (0)        |       |
| CheckMate 227 | ABCP       | 1.78 (0.72)   | 17.95 (1.85)  | Arches | Placebo      | 0 (0)         | 0 (0)         | BrightTNess  | Bev       | -3.56 (2.27)  | -5.87 (2)    |       |
| CheckMate 227 | Nivo + Ipi | 44.05 (0.67)  | 38.21 (1.38)  | Arches | Apalutamide  | 70.25 (2.66)  | 195.34 (4.74) | BrightTNess  | Carbo     | 8.94 (2.21)   | 0.89 (1.88)  |       |
| CheckMate 227 | Pembro     | 46.51 (1.97)  | 54.84 (4.22)  | Arches | Enzalutamide | 23.47 (0.74)  | 4.73 (0.9)    | BrightTNess  | P + C     | 6.47 (3.49)   | -1.29 (3.42) |       |
| IMPower150    | Chemo      | 0 (0)         | 0 (0)         | TITAN  | Placebo      | 0 (0)         | 0 (0)         | BrightTNess  | P + C + V | 17.88 (3.12)  | 20.20 (2.76) |       |
| IMPower150    | ABCP       | 2.20 (0.31)   | -60.16 (4.18) | TITAN  | Apalutamide  | 33.83 (1.43)  | 43.76 (2.29)  | BrightTNess  | Pembro    | 108.44 (1.75) | 84.45 (1.41) |       |
| IMPower150    | Nivo + Ipi | 10.99 (0.44)  | -54.13 (4.39) | TITAN  | Enzalutamide | 13.40 (0.48)  | 6.12 (0.35)   | CALGB 40603a | Placebo   | 0 (0)         | 0 (0)        |       |
| IMPower150    | Pembro     | 17.20 (0.95)  | -27.93 (6.03) |        |              |               |               | CALGB 40603a | Bev       | 5.41 (1.18)   | 6.46 (1.07)  |       |
| Keynote-189   | Chemo      | 0 (0)         | 0 (0)         |        |              |               |               | CALGB 40603a | Carbo     | 2.42 (1.25)   | 1.36 (1.16)  |       |
| Keynote-189   | ABCP       | 2.98 (0.59)   | 20.37 (1.43)  |        |              |               |               | CALGB 40603a | P + C     | 26.56 (2)     | 22.29 (2.24) |       |
| Keynote-189   | Nivo + Ipi | 26.51 (0.89)  | 28.19 (1.17)  |        |              |               |               | CALGB 40603a | P + C + V | 32.72 (1.69)  | 32.78 (1.71) |       |
| Keynote-189   | Pembro     | 29.50 (1.44)  | 66.70 (3.6)   |        |              |               |               | CALGB 40603a | Pembro    | 66.12 (0.77)  | 60.57 (0.7)  |       |
|               |            |               |               |        |              |               |               | CALGB 40603b | Placebo   | 0 (0)         | 0 (0)        |       |
|               |            |               |               |        |              |               |               | CALGB 40603b | Bev       | 2.01 (1.27)   | 3.09 (1.17)  |       |
|               |            |               |               |        |              |               |               | CALGB 40603b | Carbo     | 7.06 (1.17)   | 6.03 (1.08)  |       |
|               |            |               |               |        |              |               |               | CALGB 40603b | P + C     | 26.99 (1.99)  | 22.02 (2.31) |       |
|               |            |               |               |        |              |               |               | CALGB 40603b | P + C + V | 33.13 (1.7)   | 32.72 (1.76) |       |
|               |            |               |               |        |              |               |               | CALGB 40603b | Pembro    | 65.63 (0.75)  | 59.72 (0.69) |       |
|               |            |               |               |        |              |               |               | Keynote-522  | Placebo   | 0 (0)         | 0 (0)        |       |
|               |            |               |               |        |              |               |               | Keynote-522  | Bev       | -5.46 (2.09)  | -6.39 (1.7)  |       |
|               |            |               |               |        |              |               |               | Keynote-522  | Carbo     | 5.36 (2.01)   | -0.57 (1.62) |       |
|               |            |               |               |        |              |               |               | Keynote-522  | P + C     | 3.29 (3.2)    | -5.36 (3.3)  |       |
|               |            |               |               |        |              |               |               | Keynote-522  | P + C + V | 11.96 (2.88)  | 10.55 (2.72) |       |
|               |            |               |               |        |              |               |               | Keynote-522  | Pembro    | 100.8 (1.79)  | 77.39 (1.3)  |       |

The standard deviation is noted in brackets.  
ABCP: Atezolizumab, bevacizumab, carboplatin and paclitaxel; eBC: Early breast cancer; mPC: Metastatic prostate cancer; Nivo + Ipi: Nivolumab and ipilimumab; P+C: Paclitaxel and carboplatin; P+C+V: Paclitaxel, carboplatin and velparib.

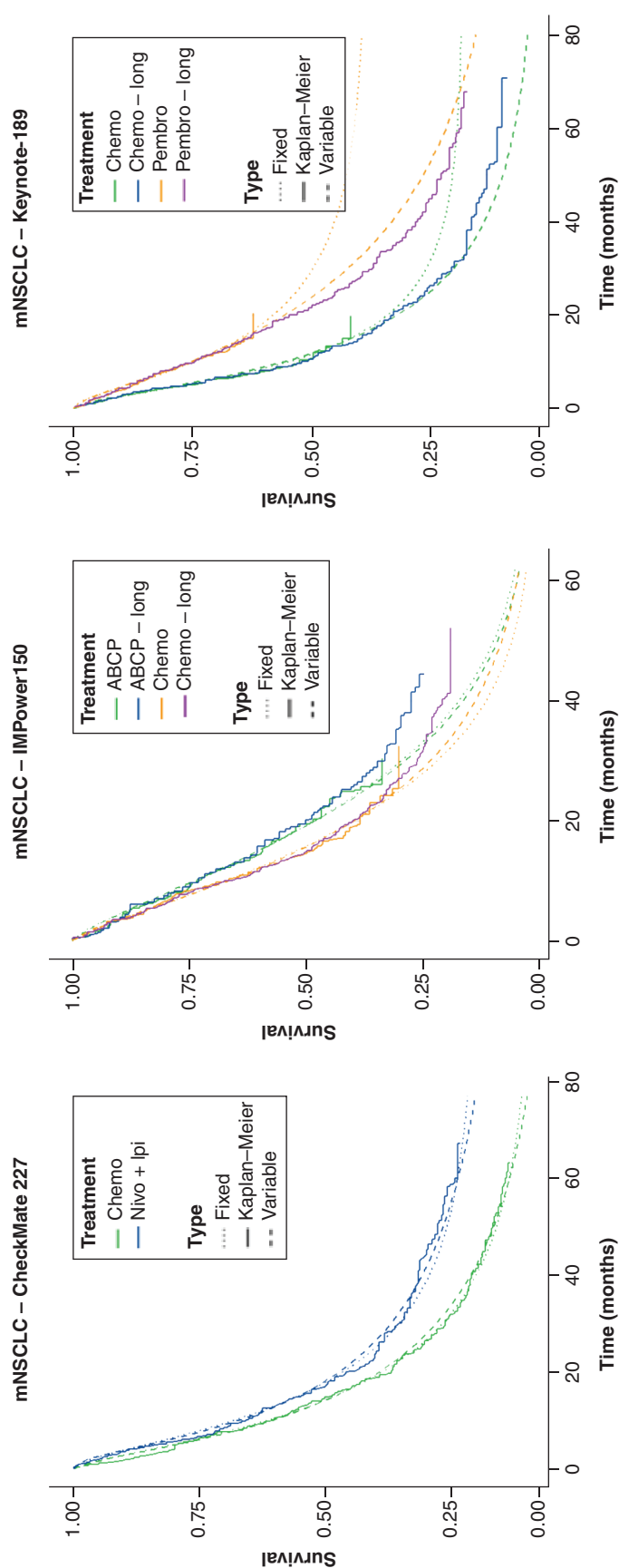


Figure 4. Visual fits of the advanced non-small-cell lung cancer network. Modeled survival curves are presented per trial, overlaid with observed Kaplan-Meier estimates. Long-term follow-up curves are shown for visual comparison only and were not used in model fitting. ABCP: Atezolizumab, bevacizumab, carboplatin and paclitaxel.

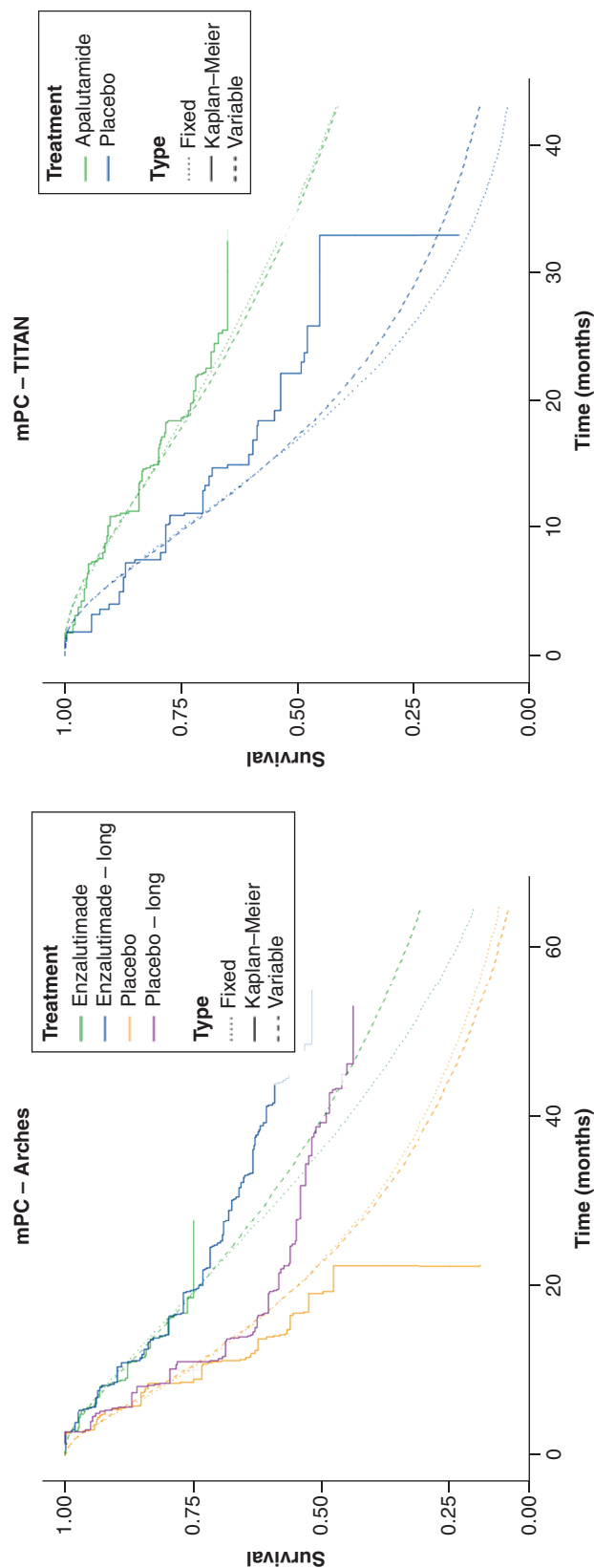
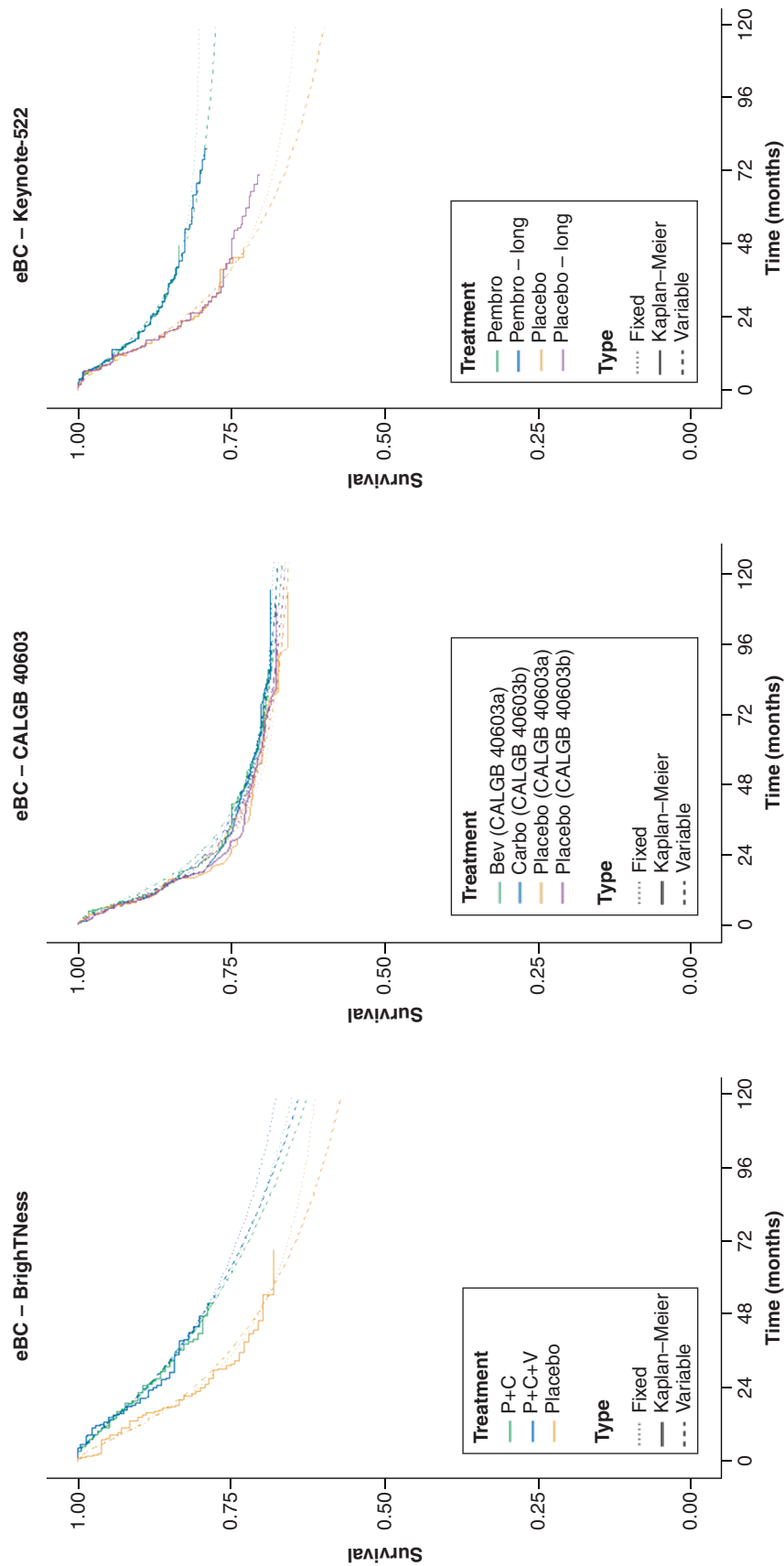


Figure 5. Visual fits of the metastatic prostate cancer network. Modeled survival curves are presented per trial, overlaid with observed Kaplan-Meier estimates. Long-term follow-up curves are shown for visual comparison only and were not used in model fitting. mPC: Metastatic prostate cancer.



**Figure 6. Visual fits of the early breast cancer network.** Modeled survival curves are presented per trial, overlaid with observed Kaplan-Meier estimates. Long-term follow-up curves are shown for visual comparison only and were not used in model fitting.  
 eBC: Early breast cancer; P+C: Paclitaxel and carboplatin; P+C+V: Paclitaxel, carboplatin and veliparib.

| Table 4. Execution time for the variable and fixed models, for each dataset.                             |                   |         |       |         |        |         |
|----------------------------------------------------------------------------------------------------------|-------------------|---------|-------|---------|--------|---------|
|                                                                                                          | eBC               |         | mPC   |         | mNSCLC |         |
|                                                                                                          | Total             | Average | Total | Average | Total  | Average |
| Variable                                                                                                 | 15.0              | 5.0     | 7.9   | 2.6     | 6.9    | 2.3     |
| Fixed                                                                                                    | 58.1 <sup>†</sup> | 2.3     | 30.4  | 1.1     | 39.3   | 1.4     |
| The execution time is in hours.                                                                          |                   |         |       |         |        |         |
| <sup>†</sup> Three fixed models did not converge, the execution time has been excluded for these models. |                   |         |       |         |        |         |
| eBC: Early breast cancer; mPC: Metastatic prostate cancer.                                               |                   |         |       |         |        |         |

both the observed long-term KM curve and the variable model projection. Similar patterns are seen in the IMpower150 arm, where the variable model’s projection for ABCP aligns more closely with the long-term KM data.

In the mPC network, visual differences in extrapolated tails are most pronounced in the TITAN and ARCHES trials, see [Figure 5](#). Notably, the placebo arms in these studies show markedly different empirical survival patterns, with TITAN exhibiting a sharper early decline. The model fits reflect this: in TITAN, the fixed model substantially overpredicts long-term survival for placebo, diverging from the observed curve, whereas the variable model more closely follows the downward trend. In contrast, in ARCHES, both models track the empirical data reasonably well over the observed period, though the variable model yields a slightly more gradual extrapolation. These patterns illustrate how underlying differences in trial populations or event dynamics can influence model behavior and underscore the value of flexible time transformations in capturing study-specific hazard trends.

In the eBC network, differences between fixed and variable models are more subtle, see [Figure 6](#). Across trials like BrighTNess and CALGB 40603, both models show comparable extrapolations. In the KEYNOTE-522 study, however, the fixed model tends to maintain higher long-term survival estimates, especially for the placebo arm, whereas the variable model exhibits a more tapered slope consistent with the extended follow-up curve.

Across all datasets, these visualizations illustrate that while both model types achieve good fits within the observed data range, their extrapolation behavior differs systematically. These differences are further quantified in the RMSE and long-term validation analyses.

Computational efficiency

[Table 4](#) summarizes the total and average execution times (in minutes) for both variable and fixed FP models across the three datasets. Although the average runtime per fixed model is lower (e.g., 2.3 h for the eBC dataset) than that of the variable models (5.0 h for eBC), the fixed approach requires running 28 distinct model configurations to cover the full grid of power combinations, versus three variable models. Consequently, the overall computational burden is substantially higher for the fixed models. For example, the total execution time for the fixed models in the eBC dataset is 58.1 h (28-times 5.3 h per model on average), compared with only 15.0 h for the variable approach. Similar trends are observed for the mPC and aNSCLC datasets, where the fixed models incur total times of 30.4 and 39.3 h, respectively, versus 7.9 and 6.9 h for the variable models. It is also noteworthy that three fixed models did not converge and were excluded from these totals. These results underscore the computational efficiency gains achieved by the variable power approach, which streamlines model selection by eliminating the need to run multiple fixed configurations.

Uncertainty analysis

To complement statistical and visual model assessments, we examined both posterior and predictive uncertainty for fixed and variable FP models.

While variable models introduce two additional parameters, the powers  $p_1$  and  $p_2$ , these do not necessarily inflate overall uncertainty. In fact, our results suggest the opposite: despite their flexibility, variable models frequently yield more concentrated posterior distributions, particularly in long-term survival projections and incremental survival estimates. This is likely due to their ability to better capture the underlying hazard structure, thereby reducing residual variance in fitted and extrapolated survival curves.

Across all datasets, the posterior distributions of  $p_1$  and  $p_2$  were unimodal and sharply peaked, indicating well-identified transformations and stable model behavior. [Supplementary Figures 2–4](#) in the Appendix illustrate these joint distributions, supporting the notion that the added degrees of freedom in the variable models are efficiently used, not arbitrarily absorbed into noise.

Importantly, when comparing uncertainty in incremental survival outcomes, variable models generally showed comparable or lower standard deviations than their fixed counterparts, despite the extra parameters. This reinforces their utility for downstream decision-making tasks, such as probabilistic sensitivity analyses in cost-effectiveness models, where both predictive accuracy and coherent uncertainty quantification are essential.

## Discussion

Our study demonstrates that variable power FP models offer notable improvements over traditional fixed power approaches for modeling survival data in an NMA context. Variable power models consistently achieved lower LOOIC values and competitive RMSE metrics across the three datasets, indicating superior predictive accuracy while maintaining an adequate fit to the observed KM data. Although both fixed and variable models produced generally similar visual fits to the empirical survival curves, the variable models yielded more plausible extrapolations and avoided the extreme or unrealistic long-term predictions sometimes observed in fixed models. This behavior translated into incremental survival estimates that were more stable and less prone to over- or underestimation.

The statistical benefits of continuously estimating the power parameters are twofold. First, the flexibility of the variable approach allows the model to capture subtle nonlinear hazard trends that might otherwise be overlooked when restricted to a limited set of pre-defined values. Second, the continuous estimation process fine-tunes the time transformation and enhances predictive performance without introducing undue uncertainty. Although variable models include two additional parameters ( $p_1$ ,  $p_2$ ), we found that their posterior distributions were tightly concentrated and unimodal (Supplementary Figures 5–7), indicating stable, well-identified solutions. This is likely due to the smooth, convex structure of the FP parameter space, which supports efficient inference in Bayesian settings.

From a clinical perspective, the plausibility of extrapolated survival curves is often judged against observed long-term trial follow-up or expert expectations. In our datasets, the variable models generally aligned more closely with available extended KM data (e.g., in IMpower150 and KEYNOTE-189), suggesting their long-term projections may better reflect realistic survival dynamics. These calibrated extrapolations help avoid over- or underestimation of incremental survival, which is critical for downstream applications such as comparative effectiveness and cost-effectiveness analyses in HTA settings.

Furthermore, the variable power approach significantly reduces the overall computational burden by eliminating the need to run an exhaustive grid of fixed power configurations. Rather than evaluating dozens of fixed models, the continuous framework allows the data to guide the estimation process, thereby streamlining model selection and reducing runtime without compromising fit. This efficiency gain is particularly relevant in high-dimensional settings where computational resources and convergence time are critical considerations. Crucially, these computational advantages are achieved without compromising model validation, as evidenced by superior predictive performance, calibrated long-term extrapolations and consistent visual and statistical fit across datasets.

Another modeling choice relates to how treatment and study effects are assigned across parameters. In our current specification, treatment and study effects were defined for all  $\beta$  parameters (scale and shape). This allowed us to focus directly on the comparison between fixed and variable powers without introducing additional degrees of freedom in model specification. However, prior work has shown that applying treatment effects to all FP terms can risk overfitting, particularly in smaller or sparser networks [13]. In practice, analysts may wish to restrict treatment effects to the scale parameter alone, or to the scale and a single shape parameter, depending on the data and the clinical question. The flexibility of our framework accommodates such specifications, and future applications may benefit from tailoring the placement of treatment and study effects to balance parsimony with plausibility.

Our findings extend previous work by Royston and Parmar, and by Jansen, by relaxing the fixed power constraint that has long been a limitation in FP modeling [2,3]. The novelty of our approach lies in the use of STAN to continuously estimate the power parameters, thereby refining the modeling process and improving both statistical and clinical outcomes, providing full code for reproducibility. In doing so, our method provides a more nuanced representation of the underlying survival dynamics, which is essential when comparing multiple treatment modalities in an NMA framework. Relatedly, Heeg *et al.* explored an implementation of variable powers in an NMA context, demonstrating the feasibility of relaxing fixed powers [12]. However, their work did not present a formal mathematical formulation, and the specification they employed differed from ours. Specifically, our formulation introduces explicit constraints to stabilize estimation, separates the parameter space into negative/negative, negative/positive and positive/positive configurations to avoid instability around zero, and enforces ordering con-

straints to prevent symmetry-related local optima. These innovations distinguish our approach, providing a more rigorous and reproducible framework for estimating variable-power FP models.

Nevertheless, several limitations warrant consideration. The current study focuses solely on the standard FP formulation, and alternative approaches, such as repeated measures FP or FP models incorporating log time transformations, were not explored. Additionally, although we have employed weakly informative priors to mitigate undue influence, our results remain potentially sensitive to prior specification and convergence challenges common in complex Bayesian models. Moreover, our reliance on pseudo-individual patient data, derived from digitized KM curves, introduces additional uncertainty compared with analyses based on true individual-level data.

This study did not incorporate general population mortality in the fitting of the FP models. While adjusting for background mortality is standard in certain health-economic settings, particularly when modeling long-term survival or when extrapolating beyond the trial population's typical life expectancy, its omission here was a deliberate choice to maintain consistency across model comparisons and reduce computational burden. Nonetheless, incorporating general population mortality is straightforward within the proposed framework. To support reproducibility and downstream use, example code demonstrating how to implement general population mortality adjustments is provided in the Appendix.

Future research should aim to expand the methodology by incorporating alternative FP formulations, including those that account for repeated measures and log time transformations. In addition, the current formulation could be extended to allow for treatment-specific effects on the power parameters themselves, enabling the model to capture variation in the shape of the hazard curve attributable to treatment. Furthermore, the consistent positioning of the optimized power parameters between the best-performing fixed configurations suggests that the underlying survival model landscape may exhibit local convexity. This observation raises the possibility that future frequentist approaches, leveraging convex optimization techniques such as the Newton–Raphson algorithm, could efficiently identify optimal transformations without exhaustive search. Further investigation is warranted into the propagation of parameter uncertainty in downstream applications, particularly cost-effectiveness analyses using probabilistic sensitivity analysis.

In conclusion, the variable power FP models not only improve statistical fit and predictive accuracy but also reduce structural uncertainty by replacing exhaustive grid-search with a unified, data-driven estimation of transformation powers. They produce clinically plausible extrapolations and more stable incremental survival estimates. These improvements can better inform comparative effectiveness research and health policy decision-making by streamlining model selection and enhancing confidence in long-term projections. Our findings underscore the value of continuous power estimation in overcoming the limitations of fixed FP models, providing a robust, flexible framework for modeling complex survival data in oncology and beyond.

### Summary points

- Fractional polynomials (FPs) are widely used for flexible survival modeling but traditionally require selecting transformation powers from a fixed grid.
- This grid-based approach can be inefficient, structurally uncertain and prone to implausible extrapolations.
- We present the first formal Bayesian formulation and Stan implementation of variable-power FP models for survival analysis in a network meta-analysis framework.
- In our approach, the FP powers are estimated directly from the data as random variables rather than treated as fixed inputs.
- To ensure stability, we introduced explicit constraints: ordering of powers, partitioning of negative/positive configurations and bounded ranges.
- Across three oncology network meta-analysis (aNSCLC, metastatic castration-sensitive prostate cancer, HER2+ early breast cancer), variable-power models consistently achieved better predictive accuracy than fixed-power models.
- Variable-power models produced more clinically plausible long-term extrapolations and more stable incremental survival estimates.
- The continuous estimation framework reduced computational burden by eliminating exhaustive grid searches of fixed-power combinations.
- Our specification differs from earlier exploratory implementations [12] by providing a full mathematical formulation and enhanced stability.
- These improvements strengthen survival extrapolation in health technology assessment and comparative effectiveness research, supporting more reliable health policy decision-making.

### Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: <https://becarispublishing.com/doi/epdf/10.57264/cer-2025-0126>

### Financial disclosure

The authors received no financial and/or material support for this research or the creation of this work.

### Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

### Writing disclosure

No funded writing assistance was utilized in the production of this manuscript.

### Ethical conduct of research

No ethics approval was required, as only published KM curves were used to reconstruct pseudo-individual patient level data.

### Data sharing statement

The STAN code is provided in the Appendix.

### Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J. R. Stat. Soc. Ser. C Appl. Stat.* 43(3), 429–453 (1994).
- **Seminal paper introducing the fractional polynomial (FP) approach for regression; cornerstone for all subsequent FP survival modeling.**
2. Royston P, Sauerbrei W. *Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables*. John Wiley & Sons (2008).
- **Landmark book detailing the FP methodology, practical implementation and influence on modern survival analysis.**
3. Jansen JP. Network meta-analysis of survival data with fractional polynomials. *BMC Med. Res. Methodol.* 11, 61 (2011).
- **First application of FP models within survival network meta-analysis (NMA); foundational for methodological extensions in health technology assessment.**
4. Rutherford MJ, Lambert PC, Sweeting MJ, Pennington R, Crowther MJ, Abrams KR. *NICE DSU Technical Support Document 21: Flexible methods for survival analysis*. Decision Support Unit, ScHARR, University of Sheffield, UK (2020). Available from: <https://sheffield.ac.uk/nice-dsu/tsds/flexible-methods-survival-analysis>
5. NICE Technology Appraisal Guidance 836. Palbociclib with fulvestrant for treating hormone receptor-positive, HER2-negative advanced breast cancer after endocrine therapy (2022). Available from: <https://www.nice.org.uk/guidance/ta836>
6. NICE Technology Appraisal Guidance 584. Atezolizumab in combination for treating metastatic non-squamous non-small-cell lung cancer (2019). Available from: <https://www.nice.org.uk/guidance/ta584>
7. NICE Technology Appraisal Guidance 463. Cabozantinib for previously treated advanced renal cell carcinoma (2017). Available from: <https://www.nice.org.uk/guidance/ta463>
8. NICE Technology Appraisal Guidance 964. Cabozantinib with nivolumab for untreated advanced renal cell carcinoma (2024). Available from: <https://www.nice.org.uk/guidance/ta964>
9. Leahy T, Besford M, Sammon C. PMU109 how do NICE evidence review groups and appraisal committees perceive fractional polynomial methodology in technology appraisals? *Value Health* 22(Suppl. 3), S727 (2019).
10. Verhoek A, Ouwens M, Heeg B. MSRI34 optimizing fractional polynomials by using variable powers. *Value Health* 25(Suppl. 12), S376 (2022).
11. Plummer M. JAGS Version 4.3.0 User Manual (2017). Available from: [http://people.stat.sc.edu/hansont/stat740/jags\\_user\\_manual.pdf](http://people.stat.sc.edu/hansont/stat740/jags_user_manual.pdf)

12. Heeg B, Garcia A, Beekhuizen SV *et al.* Novel and existing flexible survival methods for network meta-analyses. *J. Comp. Eff. Res.* 11(15), 1121–1133 (2022).
- **First published application exploring variable powers in NMA; demonstrates feasibility but differs in formulation from the current approach.**
13. Freigofaite D, Verhoek A, Heeg B. POSA303 Influence of treatment effect on parameters in fractional polynomials for network meta-analyses. *Value Health* 25(Suppl. 1), S198–S199 (2022).
14. Statistics Netherlands. 2025. (Accessed: 30 May 2025). Available from: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37360ned/table>
15. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford University Press, UK (2015).
16. Sauerbrei W, Royston P, Look M. A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biom. J.* 49(3), 453–473 (2007).
- **Extends FP methodology to time-varying effects; illustrates the versatility and continuing development of FP approaches.**
17. Betancourt M. How the shape of a weakly informative prior affects inferences. Stan User's Guide March. (2017). Available from: [http://people.stat.sc.edu/hansont/stat740/jags\\_user\\_manual.pdf](http://people.stat.sc.edu/hansont/stat740/jags_user_manual.pdf)
18. Mitchel M, Muftakhidinov B, Winchen T *et al.* markummittchell/engage-digitizer: Nonrelease (v12.2.1). Zenodo. (2020). <https://doi.org/10.5281/zenodo.3941227>
19. Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med. Res. Methodol.* 12, 9 (2012).
20. O'Byrne K, Popoff E, Badin F, Lee A, Yuan Y, Lozano-Ortega G *et al.* Long-term comparative efficacy and safety of nivolumab plus ipilimumab relative to other first-line therapies for advanced non-small-cell lung cancer: a systematic literature review and network meta-analysis. *Lung Cancer* 177, 11–20 (2023).
21. Socinski MA, Nishio M, Jotte RM *et al.* Impower150 final overall survival analyses for atezolizumab plus bevacizumab and chemotherapy in first-line metastatic nonsquamous nscl. *J. Thoracic Oncol.* 16(11), 1909–1924 (2021).
22. Garassino MC, Gadgeel S, Speranza G *et al.* Pembrolizumab plus pemetrexed and platinum in nonsquamous non-small-cell lung cancer: 5-year outcomes from the Phase 3 Keynote-189 study. *J. Clin. Oncol.* 41(11), 1992–1998 (2023).
23. Wang L, Paller CJ, Hong H, De Felice A, Alexander GC, Brawley O. Comparison of systemic treatments for metastatic castration-sensitive prostate cancer: a systematic review and network meta-analysis. *JAMA Oncol.* 7(3), 412–420 (2021).
24. Armstrong AJ, Shore ND, Szmulewitz RZ *et al.* Efficacy of enzalutamide plus androgen deprivation therapy in metastatic hormone-sensitive prostate cancer by pattern of metastatic spread: ARCHES *post hoc* analyses. *J. Urol.* 205(5), 1361–1371 (2021).
25. Wilson FR, Coombes ME, Wylie Q *et al.* Herceptin® (trastuzumab) in HER2-positive early breast cancer: protocol for a systematic review and cumulative network meta-analysis. *Syst. Rev.* 6(1), 196 (2017).
26. Koornstra R, Hernandez-Aya L, Di Giacomo AM *et al.* Longer follow-up confirms recurrence-free survival benefit of adjuvant pembrolizumab in high-risk stage III melanoma: updated results from the EORTC 1325-mg/Keynote-054 trial. *J. Clin. Oncol.* 38(33), 3925–3936 (2020).
27. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* 27, 1413–1432 (2017).