Functional Cox Regression for Time-to-Event Prediction: Learning Phase-Specific Effects from Longitudinal Trajectories

Abstract

Time-to-event prediction is central in applications where decisions depend not only on *whether* an event will occur but also on *when*. In modern settings, individuals are often monitored longitudinally, yielding irregular trajectories of performance, biomarkers or sensor readings. Standard survival models typically compress these trajectories into scalar summaries, sacrificing information about *which* phases are most predictive. Black-box methods can exploit rich features but rarely provide phase-specific interpretation.

We apply and empirically evaluate a functional Cox regression framework that treats each trajectory as a functional covariate. The log-hazard includes a linear functional of the trajectory over a normalised phase variable, yielding a coefficient function that reveals how different parts contribute to the hazard. We compare four model classes: scalar Cox proportional hazards, random survival forest, functional linear Cox model (FLCM) and additive functional Cox model (AFCM) with nonlinear effects.

As a case study we analyse annual season-best performances (one observation per athlete–event–year) from 67,977 athlete–event careers of Olympic-level track-and-field athletes, yielding 347,625 yearly observations extracted from over 1.3 million competitive performances. The goal is to predict time from event debut to peak performance, which has direct implications for talent identification, resource allocation in national sporting programs, and athlete-specific career planning. Across event families, functional models consistently achieve the best discrimination and rank correlation, showing that trajectory shape contains substantial information beyond scalar summaries. After calibration, functional models also attain lower prediction errors in most families, although scalar Cox remains competitive.

The estimated coefficient $\beta(s)$ can be summarised over early, mid and late career phases, highlighting when above-typical performance predicts earlier or later peaks. These phase-specific

effects cannot be extracted from scalar models and provide actionable insights for performance forecasting and intervention timing. We argue that functional Cox regression offers a practical and interpretable approach for time-to-event prediction from longitudinal trajectories in domains ranging from sport science to medicine, reliability and behavioural analytics.

Keywords: Functional Data Analysis; Explainable AI; Longitudinal Phenotyping; Sports Analytics; Survival Analysis.

1   Introduction

Time-to-event prediction problems arise whenever we wish to forecast *when* a particular outcome will occur. Examples include time-to-disease onset or progression in medicine, time-to-equipment failure in reliability, time-to-user churn in digital platforms, and time-to-educational milestones. In many of these applications an individual's state is monitored repeatedly before the event, generating rich longitudinal trajectories of biomarkers, performance scores, usage patterns or sensor traces.

Classical survival analysis methods typically incorporate such data through a small number of engineered summaries—for example, the current value of a biomarker, a slope estimate or an accumulated exposure. This strategy is simple and often effective, but it obscures the *temporal structure* of the trajectory and makes it impossible to answer questions such as: *which phase of the trajectory is most predictive for the event time?* Black-box machine learning approaches such as random survival forests or deep survival networks can ingest high-dimensional features but usually provide limited insight into phase-specific effects.

Functional data analysis (FDA) offers an alternative perspective in which longitudinal trajectories are treated as smooth functions rather than as collections of summary statistics [Ramsay and Silverman, 2005]. In a functional regression framework one relates scalar or functional responses to functional predictors via coefficient functions that vary over the trajectory domain. When combined with Cox-type survival models, this leads to *functional Cox regression*, where the log-hazard includes a term of the form $\int_0^1 \beta(s)X_i(s)\,ds$ involving an unknown coefficient function $\beta(s)$ and an individual trajectory $X_i(s)$. Regions of the domain where $|\beta(s)|$ is large correspond to phases in which the trajectory carries strong information about the hazard.

This functional Cox framework addresses both limitations noted above: by modelling the full

trajectory $X_i(s)$ it retains phase-specific information that scalar summaries discard, while the explicit coefficient function $\beta(s)$ provides interpretable insight into *which* phases drive predictions— something opaque in black-box models. Moreover, with explicit smoothness assumptions and a well-defined functional form, functional Cox models can achieve stable estimates with moderate sample sizes (tens of thousands rather than millions), making them particularly well-suited for settings where interpretability and sample efficiency are valued alongside predictive performance.

In this paper we apply and evaluate a functional Cox regression framework for time-to-event prediction from longitudinal trajectories. Our focus is methodological: we wish to understand when modelling full trajectories as functional covariates yields tangible gains over strong scalar baselines, and what kinds of additional insight functional models can provide. To ground the discussion we use a large-scale case study from sport science: predicting the time from event debut to performance peak in elite track-and-field athletes. Here each individual is observed through an irregular performance trajectory over their career, and the event of interest is the year of peak performance in a given event. This setting provides a challenging, data-rich example of a general problem: predicting an event time from an informative longitudinal trajectory.

Using a database of annual season-best results (extracted from over one million competitive performances), we construct standardized performance trajectories for each athlete–event combination and define time-to-peak as the number of years from first recorded season to the (possibly censored) year of peak performance. We then compare four classes of models within broad event families: (i) a scalar Cox proportional hazards model using career length and observation count as covariates; (ii) a random survival forest using the same scalars; (iii) a functional linear Cox model (FLCM) in which the log-hazard depends linearly on the full performance trajectory; and (iv) an additive functional Cox model (AFCM) allowing nonlinear functional effects.

Our main contributions are as follows:

(i) We present a practical pipeline for functional Cox regression with irregular longitudinal trajectories, including time normalization, spline smoothing, within-career standardisation and a simple calibration step that maps model risk scores to absolute event times.

(ii) We provide a unified empirical comparison of scalar Cox, random survival forest, FLCM and AFCM models across multiple event families, using harmonised covariates and a common

evaluation protocol based on both hazard-based discrimination and calibrated time-to-event prediction.

(iii) Across families, the additive functional Cox model (AFCM) achieves the best overall performance in C-index (hazard-based discrimination) compared to the functional linear, scalar Cox, and random forest models. Furthermore, after calibration, the AFCM also yields the lowest Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for predicting the time-to-peak, demonstrating that nonlinear functional effects are crucial for accurate absolute time prediction. The functional linear Cox model (FLCM) remains a strong second-best performer and provides a transparent coefficient function for phase-specific interpretability.

(iv) Through the case study on peak performance in track-and-field, we illustrate how these phase-specific effects align with substantive notions such as "early peakers" and "late developers", and discuss how similar insights could be exploited in other application domains.

The rest of the paper is organized as follows. Section 2 describes the dataset, outcome definition, and construction of functional trajectories. Section 3 introduces the scalar and functional survival models, along with the evaluation framework. Section 4 presents empirical results across event families. Section 5 interprets the estimated functional effects and visualizes phase-specific patterns. Section 6 discusses implications and broader applications, and Section 7 concludes.

2   Data and Problem Formulation

2.1   Dataset and event families

We analyze a large-scale dataset of track-and-field performances from athletes who competed at Olympic Games between 1996 and 2024. For each Olympic athlete, we obtained their complete competitive career history via the World Athletics public API (worldathletics.org), yielding annual season-best results for male and female athletes across 93 distinct events. The data span each athlete's full career from debut through early 2025, ensuring comprehensive longitudinal trajectories for world-class careers. After preprocessing, the working dataset contains $N_{\mathrm{AE}} = 67,977$ athlete–event combinations and 347,625 yearly observations.[1]

---

[1]These counts are obtained after applying all data quality filters and inclusion criteria described in Section 2.3. The exact figures are reproducible by running the accompanying R code (see supplementary materials).

Events are grouped into broad families reflecting physiological and biomechanical similarities:

- **Sprint**: short track events (e.g., 100m, 200m, 400m).

- **Hurdles**: short and intermediate hurdle events.

- **Middle distance**: track events in the approximate 800–3000m range.

- **Long distance**: longer track events (e.g., 5000m, 10 000m) and track-based distance races.

- **Road**: road races (e.g., 10km, half-marathon, marathon).

- **Field**: jumps and throws. Although these disciplines differ biomechanically, we group them together as non-cyclical, technique-intensive power events that contrast with the cyclical running disciplines. Empirical analysis confirms that functional models achieve strong discrimination within this combined category, suggesting that the flexible baseline hazard and trajectory-specific functional effects adequately capture heterogeneity between jumping and throwing events.

- **Combined**: multi-event competitions (e.g., decathlon, heptathlon).

Within each family, we fit separate models that share the same covariate specification but allow the baseline hazard and regression parameters to differ.

2.2 Standardized performance metric

Raw performance measures (times, distances, points) are not directly comparable across events, sexes, or eras. We therefore use the standardized performance score provided by the World Athletics database (`result_score`), which maps each competitive result to a dimensionless score on which larger values consistently correspond to better performance. This official scoring system accounts for event-specific scales, historical performance distributions, and ensures comparability across different disciplines and competition eras. The resulting score reflects a performance level relative to the broader field of athletes competing in the same event.

For each athlete–event combination, we extract the standardized annual season-best score (the maximum `result_score` achieved in each calendar year) for each year in which the athlete records

a result in that event. This yields an irregularly spaced longitudinal series for each athlete–event pair, which we then embed on a common time grid as described below.

Why annual season-bests?   Although the underlying database contains multiple performances per athlete within a calendar year, we summarize each athlete–event–year by the standardized *annual season-best* and treat $\{Y_{ij}\}_{j=1}^{L_i}$ as the longitudinal series used to construct the functional trajectory $X_i(s)$. This design aligns the longitudinal resolution with our event definition (the *year* of peak performance) and improves comparability across athletes whose within-year competition schedules are highly irregular. Using all within-year results to fit $X_i(s)$ would in principle provide denser trajectories, but it introduces several practical complications: athletes who compete more frequently would contribute disproportionately many observations; the timing and frequency of competitions are likely informative (injury, qualification attempts, travel, selection), creating an observation process that is not missing-at-random; and intra-year seasonality (tapering, championship peaks) would require an additional modelling layer to avoid conflating short-term peaks with long-term progression. We therefore use annual season-bests as a robust, interpretable proxy for the latent year-to-year progression curve, while retaining overall *observation density* through the scalar covariate $\log(n_{\mathrm{obs},i})$ (number of observed seasons) included in all models.

2.3   Outcome definition and censoring

Let i index athlete–event combinations. Define t_i^{first} and t_i^{last} as the first and last calendar years in which athlete i records a standardized season-best in the given event. We define the *career length* as

$$L_i = t_i^{\mathrm{last}} - t_i^{\mathrm{first}} + 1,$$

and denote by Y_{ij} the standardized season-best score in career year $j \in \{1, \ldots, L_i\}$, where $j = 1$ corresponds to calendar year t_i^{first}.

For each athlete–event i, we identify the *observed peak* as the year within the observed trajectory at which Y_{ij} attains its maximum. Let $j_i^\star$ denote the index of this peak, and define the corresponding *time to observed peak* (in years since event debut) as

$$U_i = j_i^\star - 1.$$

6

If the athlete's last observed year t_i^{last} occurs in or after 2023, the athlete is considered still active, and the observed peak U_i represents only the *best performance to date*, not necessarily the true lifetime career peak. We therefore treat such careers as right-censored. Specifically, we define the event indicator

$$\Delta_i = \begin{cases} 1, & \text{if } t_i^{\text{last}} < 2023 \quad \text{(career ended; peak observed)}, \\[2mm] 0, & \text{if } t_i^{\text{last}} \geq 2023 \quad \text{(still active; right-censored)}, \end{cases}$$
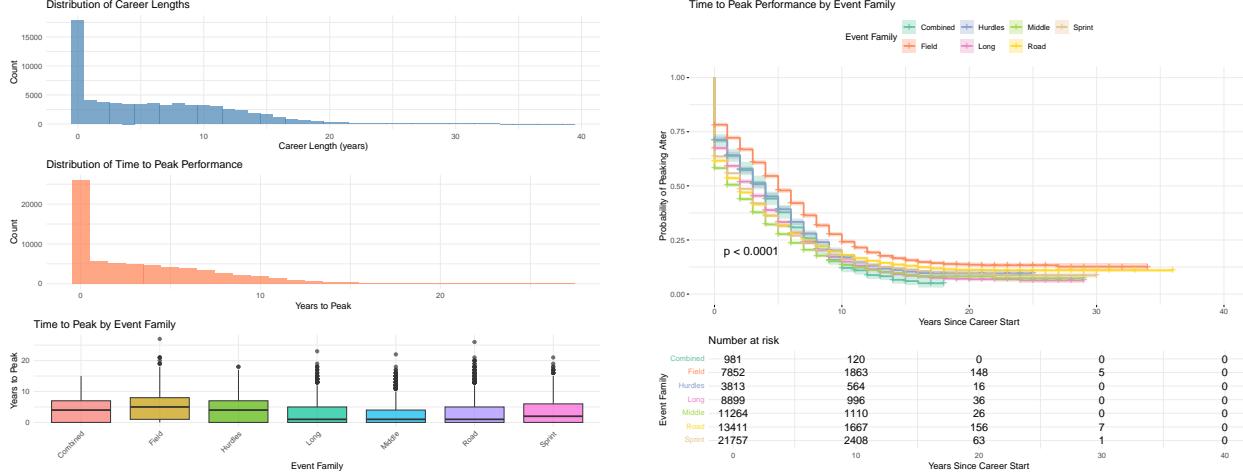
and the observed time-to-event

$$T_i = U_i.$$

For uncensored athletes ($\Delta_i = 1$), T_i represents the true time from debut to career peak. For censored athletes ($\Delta_i = 0$), T_i is the time to the best performance observed so far, but the true peak may occur in the future. Formal inclusion criteria exclude pathological cases where the trajectory does not exhibit a meaningful rise.

Across all event families, approximately 80% of athlete–event combinations have retired and exhibit an observed peak within the follow-up window ($\Delta_i = 1$), while the remaining 20% are still active and thus right censored. The median time-to-peak is 3 years from event debut (95% CI: [3, 3]), with variation across event families.

Figure 1 presents exploratory visualizations of the timing of peak performance. Panel (a) summarizes the empirical distribution of career length and time to peak. Most careers are short and both career length and time-to-peak are heavily right-skewed: the majority of athletes reach their observed peak within the first few years in an event, whereas a small subset exhibit very long careers and late peaks. Stratifying by event family reveals systematic differences, with road and long-distance events tending to peak later on average than sprints and hurdles. Panel (b) presents Kaplan–Meier estimates of the survivor function for time-to-peak by event family. The curves reveal pronounced differences in the timing of peak performance: sprinters tend to reach their peak earliest, with a steep initial decline in the probability of peaking after a given time point, whereas road and long-distance athletes maintain a higher probability of peaking later. A global log-rank test strongly rejects equality of the survivor functions ($p < 0.0001$).

(a) Distributions of career length and time to peak performance. Top: histogram of career length (years). Middle: histogram of time to peak. Bottom: boxplots by event family.

(b) Kaplan–Meier estimates of the survivor function for time to peak, stratified by event family. Log-rank test $p < 0.0001$.

Figure 1: Exploratory analysis of career trajectories and time to peak performance. Panel (a) shows distributions of career length and time to peak across all athlete–event combinations and by event family. Panel (b) shows survival curves for time to peak, revealing significant heterogeneity across event families.

2.4 Functional performance trajectories

The raw annual sequence $\{Y_{ij}\}_{j=1}^{L_i}$ represents a discrete sample of an underlying smooth performance trajectory over the course of the athlete's career in a given event. To enable functional regression, we map each athlete–event trajectory onto a common domain and obtain a smooth curve.

First, we define a normalized *career phase* variable

$$s_{ij} = \frac{j-1}{L_i - 1} \in [0, 1],$$

so that $s = 0$ corresponds to the first season and $s = 1$ to the last observed season within the career. This phase normalization is a key modeling assumption: we treat career progression as phase-relative rather than absolute-time-based, enabling comparison of trajectories across athletes with different career lengths. Under this framework, an athlete's second year (out of five) is treated as comparable to another athlete's fourth year (out of ten), since both occur at $s = 0.25$ in their respective careers. We then approximate the underlying performance trajectory $X_i(s)$ by fitting a smooth function through the points (s_{ij}, Y_{ij}). In practice we use cubic smoothing splines with

a moderate smoothing parameter to balance fidelity and smoothness. The resulting smooth curve $X_i(s)$ is evaluated on a dense, fixed grid $0 = s_1 < s_2 < \cdots < s_G = 1$ (e.g., $G = 50$), producing a vector

$$\boldsymbol{X}_i = \big(X_i(s_1), \ldots, X_i(s_G)\big)^\top.$$

To illustrate the resulting functional covariates, Figure 2 plots smoothed standardized trajectories for a random subset of 30 careers in each event family over normalized career phase $s \in [0, 1]$. Within families, trajectories share a common qualitative pattern: an initial period of improvement followed by a plateau around mid-career and, for many athletes, a decline towards the end of the observed career. The overlaid mean curves highlight systematic differences between families; for example, road and long-distance events tend to exhibit more gradual rises and longer plateaus than sprint and hurdle events. These patterns motivate the use of functional covariates to capture trajectory shape beyond simple scalar summaries.
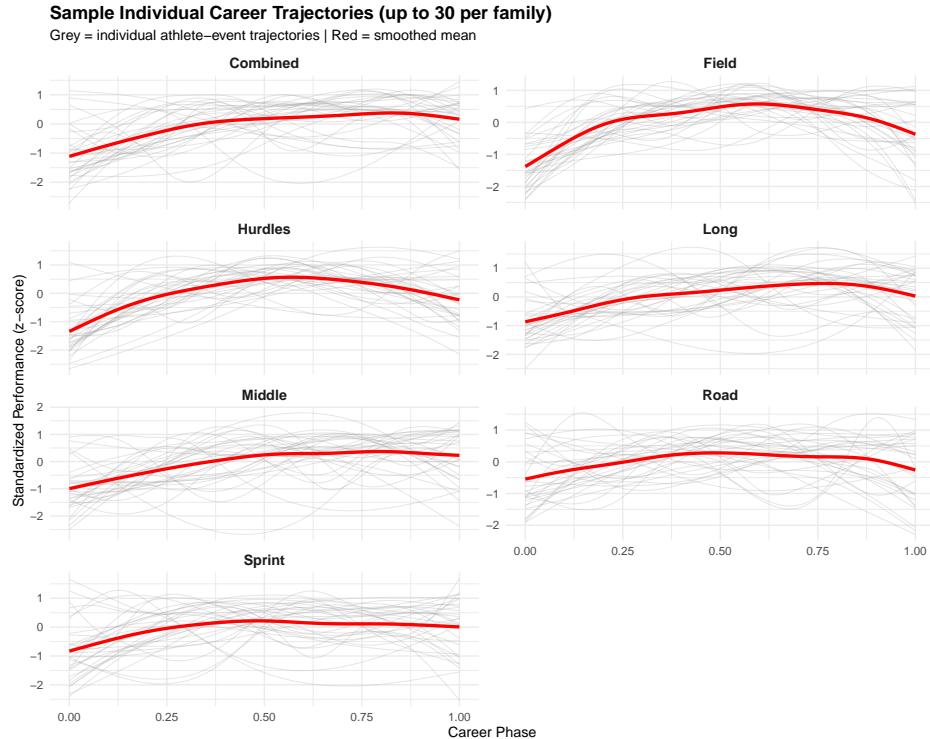

Figure 2: Sample standardized performance trajectories by event family. For each family, 30 athlete–event careers are displayed. The red curve represents the smoothed pointwise mean trajectory, obtained via generalized additive model (GAM) smoothing to highlight the population-level trend without overfitting individual noise. Trajectories are smoothed using penalized regression splines to capture the underlying progression signal while filtering out year-to-year noise. Typical patterns involve early improvement, a mid-career plateau, and eventual decline.

To stabilize functional estimation and increase comparability across athletes, we perform a within-career standardization of the trajectory, centering each $X_i(s)$ to have mean zero over $s \in [0, 1]$ and scaling by its standard deviation. This ensures that functional effects primarily capture shape differences rather than absolute scale, while the scalar variables (Section 2.5) retain information about career length and observation density.

2.5   Scalar covariates

In addition to the functional trajectory, we include two scalar covariates that summarize basic features of the career in each event:

- **Career length**: L_i as defined above. We work with a standardized version $Z_{1i} = \text{career\_length\_scaled}_i$, obtained by centering and scaling L_i within each event family.

- **Observation count**: $n_{\text{obs},i}$, the number of distinct yearly observations in the athlete–event trajectory. Since $n_{\text{obs},i}$ is highly correlated with L_i, we use a log transformation $Z_{2i} = \log(n_{\text{obs},i})$.

These scalars capture the overall duration and data density of each athlete's career in a given event. They are included in all models (scalar and functional) to ensure a fair comparison and to isolate the incremental value of the functional trajectories.

3   Modelling Framework

For each event family, we fit four classes of time-to-event models conditional on the scalar covariates Z_i and, when applicable, the functional trajectory $X_i(\cdot)$. We assume non-informative right-censoring throughout.

We assessed the proportional hazards (PH) assumption using scaled Schoenfeld residuals. A pooled Cox model including event-family indicators showed clear evidence of non-proportionality for the family effects and career-length covariate. Consequently, all Cox and functional Cox models are fitted conditional on the scalar covariates Z_i and, when applicable, the functional trajectory $X_i(\cdot)$. All parameters are estimated *separately within each family*; specifically, we estimate a family-specific functional coefficient $\beta_g(s)$ (for FLCM) or surface $f_g(s, x)$ (for AFCM). In the notation

below we suppress the family index g for simplicity, but it is implied that distinct models are fitted for Sprint, Hurdles, Middle distance, Long distance, Road, Field, and Combined. Comparison of the PH assumption for distinct families did not reveal any major systematic departures over the time range considered.

3.1 Scalar Cox proportional hazards model

The primary scalar baseline model is a Cox proportional hazards model [Cox, 1972] with hazard

$$\lambda_i(t \mid Z_i) = \lambda_0(t) \exp\left\{\boldsymbol{\gamma}^\top Z_i\right\}, \tag{1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function common to all athletes in an event family, and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^\top$ is a vector of log-hazard ratios. Within each family, we fit (1) using the standard partial likelihood approach.

This model captures how the timing of peak performance depends on simple scalar summaries of the career. For example, a negative γ_1 would indicate that athletes with longer careers (relative to their family peers) tend to reach their peak later.

3.2 Random survival forest

As a nonlinear, tree-based benchmark, we fit a random survival forest [Ishwaran et al., 2008] (RSF) using the same scalar covariates Z_i. An RSF is an ensemble of survival trees, each grown on a bootstrap sample of the data with random subsampling of covariates at each split. Each terminal node corresponds to a subgroup of athletes with similar covariates, and the node-specific survival function is estimated via the Nelson–Aalen or Kaplan–Meier estimator based on the observed event times and censoring indicators.

For each athlete i, the RSF produces an estimated cumulative hazard function $\widehat{\Lambda}_i^{\mathrm{RSF}}(t)$, obtained by averaging node-specific hazard estimates across trees. We summarize this function through a scalar *risk score* by evaluating the cumulative hazard at a fixed time point (we use the 75th percentile of observed event times): $r_i^{\mathrm{RSF}} = \widehat{\Lambda}_i^{\mathrm{RSF}}(t_{0.75})$. The RSF thus serves as a flexible, model-free comparator on the same scalar covariate set.

3.3 Functional linear Cox model (FLCM)

To incorporate the full shape of the performance trajectory, we extend the Cox model by adding a functional linear term [Kong et al., 2018]. The hazard for athlete i is specified as

$$\lambda_i(t \mid Z_i, X_i) = \lambda_0(t) \exp\left\{ \boldsymbol{\gamma}^\top Z_i + \int_0^1 \beta(s) X_i(s)\, ds \right\}, \tag{2}$$

where $\beta(s)$ is an unknown coefficient function defined on the normalized career phase $s \in [0, 1]$. The integral $\int_0^1 \beta(s) X_i(s) ds$ summarizes how the shape of $X_i(\cdot)$ over the career influences the log-hazard.

In practice, we approximate $X_i(s)$ and $\beta(s)$ using a common basis expansion. Let $\{\phi_k(s)\}_{k=1}^K$ denote a set of spline basis functions (e.g., cubic B-splines), and write

$$X_i(s) \approx \sum_{k=1}^K x_{ik}\, \phi_k(s), \qquad \beta(s) \approx \sum_{k=1}^K b_k\, \phi_k(s).$$

Then the functional term can be written as

$$\int_0^1 \beta(s) X_i(s)\, ds \approx \sum_{k=1}^K \sum_{\ell=1}^K b_k x_{i\ell} \int_0^1 \phi_k(s)\phi_\ell(s)\, ds = \boldsymbol{b}^\top \boldsymbol{M} \boldsymbol{x}_i,$$

where $\boldsymbol{M}$ is the $K \times K$ matrix with entries $M_{k\ell} = \int_0^1 \phi_k(s)\phi_\ell(s)ds$ and $\boldsymbol{x}_i$ collects the basis coefficients of X_i. Equivalently, we may work with a reduced set of functional principal component scores extracted from $\boldsymbol{X}_i$, and treat those scores as additional linear covariates.

The functional coefficient vector $\boldsymbol{b}$ and scalar coefficients $\boldsymbol{\gamma}$ are estimated by maximizing a penalized partial likelihood with roughness penalties (e.g., $\int_0^1 \{\beta''(s)\}^2 ds$) to obtain a smooth $\widehat{\beta}(s)$. The smoothing parameter is selected via REML/GCV as implemented in the `mgcv` package [Wood, 2011], and can be visualized to interpret which phases of the career are most predictive of earlier or later peak times.

3.4 Additive functional Cox model (AFCM)

The linear functional term in (2) may be overly restrictive if the effect of performance at a given career phase is nonlinear in $X_i(s)$. We therefore also consider an additive functional Cox model:

$$\lambda_i(t \mid Z_i, X_i) = \lambda_0(t) \exp\left\{\gamma^\top Z_i + \int_0^1 f\big(s, X_i(s)\big)\, ds\right\}, \tag{3}$$

where $f(s, x)$ is an unknown smooth bivariate function. In this specification, the contribution to the log-hazard at phase s depends flexibly on both s and the local performance level x. A natural implementation uses tensor-product splines:

$$f(s, x) \approx \sum_{k=1}^{K_s} \sum_{\ell=1}^{K_x} \theta_{k\ell}\, \psi_k(s)\, \phi_\ell(x),$$

with appropriate smoothing penalties over the tensor-product surface. Estimation proceeds via penalized partial likelihood, analogous to generalized additive models with a Cox likelihood.

The fitted surface $\widehat{f}(s, x)$ can be visualized as a heatmap over (s, x), providing richer insight into how performance early, mid-career, or late in the observed trajectory influences peak timing.

3.5 Retrospective prediction and cross-validation

We emphasize that this analysis is *retrospective*: we utilize the full observed trajectory $X_i(s)$ (normalized to $[0, 1]$) to characterize the relationship between career shape and peak timing. This defines a longitudinal phenotyping task rather than online forecasting.

We split the data within each event family into a training set and an independent hold-out test set, stratified at the athlete level to avoid leakage between splits. On the training set we perform 5-fold cross-validation to tune any hyperparameters (e.g., basis dimension and smoothing penalties for functional models; number of trees and node size for the RSF). Within each fold, we fit all four models (Cox, RSF, FLCM, AFCM) using the same training indices and compute risk scores on the held-out fold.

Cross-validation serves two purposes. First, it provides an internal estimate of model discrimination via concordance indices (Section 3.6.1). Second, it allows us to select hyperparameters that balance model flexibility with overfitting in a systematic way. Once tuning is complete, we refit each

model on the full training set with chosen hyperparameters and evaluate predictive performance on the hold-out test set.

3.6 Evaluation metrics

We evaluate models along two complementary dimensions: (i) their ability to correctly *rank* athletes by the timing of their peak; and (ii) their ability to produce well-calibrated *absolute* time-to-peak predictions.

3.6.1 Hazard-based evaluation

Each fitted model produces a risk score r_i for each athlete–event i in the test set. A higher risk score indicates that the model predicts the athlete will reach their peak sooner rather than later. For the scalar Cox model, $r_i = \widehat{\gamma}^\top Z_i$. For the FLCM, we set

$$r_i = \widehat{\gamma}^\top Z_i + \int_0^1 \widehat{\beta}(s) X_i(s) ds.$$

For the AFCM, we use the linear predictor $r_i = \widehat{\gamma}^\top Z_i + \int_0^1 \widehat{f}(s, X_i(s)) ds$, where $\widehat{f}(s, x)$ is the fitted bivariate function. For the RSF, we use $r_i = \widehat{\Lambda}_i^{\mathrm{RSF}}(t_{0.75})$ as described in Section 3.2.

We assess how well models rank athletes by their peak timing using Harrell's concordance index [Harrell et al., 1982] C. This measures the probability that, for any two athletes, the model correctly identifies which athlete peaked earlier. For athletes still competing (censored observations), we only compare pairs where the timing relationship is unambiguous: specifically, a pair is comparable if either both athletes have retired and reached their observed peak, or if one is still active but the other reached their peak before the first athlete's last observed competition. We also compute Spearman's rank correlation between the risk scores $\{r_i\}$ and observed times-to-peak $\{T_i\}$ among retired athletes as an additional ranking-based summary.

3.6.2 Time-to-peak prediction and calibration

While hazard-based metrics capture how well models rank athletes in terms of peak timing, practical applications often require explicit predictions of the time-to-peak for each athlete. In principle, survival models provide an estimate of the individual survival function $S_i(t)$ from which one can

extract, for example, the median time-to-peak $\tilde{T}_i$. However, in our empirical analysis we find that functional Cox models, in particular, can produce survival curves that are substantially miscalibrated in absolute time, despite good ranking performance.

To address this, we adopt a two-step calibration approach. For each event family and each model, we fit a flexible generalized additive model (GAM) that maps the model's risk score r_i to the observed time-to-peak T_i among uncensored athletes in the training set:

$$T_i = g(r_i) + \varepsilon_i,$$

where $g(\cdot)$ is a smooth function estimated via penalized splines and ε_i is an error term. The fitted function $\widehat{g}$ provides a calibrated mapping from risk scores to predicted event times. On the test set, we obtain predicted times $\widehat{T}_i = \widehat{g}(r_i^{\text{test}})$, where r_i^{test} is the risk score computed on test-set covariates. This approach recalibrates the scale of predictions while preserving the relative ranking captured by the original risk scores.

We quantify calibration-adjusted prediction accuracy using the root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{unc}}} \sum_{i:\,\Delta_i=1} (T_i - \widehat{T}_i)^2},$$

and the mean absolute error (MAE)

$$\text{MAE} = \frac{1}{N_{\text{unc}}} \sum_{i:\,\Delta_i=1} |T_i - \widehat{T}_i|,$$

where N_{unc} is the number of uncensored athletes in the test set, and Δ_i is the event indicator defined in Section 2.3.

4   Model evaluation

We evaluate models along two complementary dimensions: (i) their ability to correctly *rank* athletes by the timing of their peak; and (ii) their ability to produce well–calibrated absolute time-to-peak predictions on a held-out test set. All reported metrics are computed separately within each event family using the 20% test split; numerical values underlying the plots are provided in the

15

Supplementary Material.

4.1 Hazard-based discrimination

Each fitted model yields a scalar risk score r_i for each athlete–event i in the test set, where larger values indicate an increased hazard of reaching the peak earlier. For the scalar Cox model we use $r_i = \widehat{\gamma}^\top Z_i$, while for the FLCM we set

$$r_i = \widehat{\gamma}^\top Z_i + \int_0^1 \widehat{\beta}(s)\, X_i(s)\, ds,$$

with analogous linear predictors extracted from the AFCM and RSF. We then assess discrimination using Harrell's concordance index (C-index) and the Spearman rank correlation between $\{r_i\}$ and the observed times-to-peak $\{T_i\}$ among uncensored athletes.

Figure 3 (top panels) summarizes the hazard-based metrics. In every family the FLCM and AFCM attain the highest C-indices and rank correlations, consistently outperforming the scalar Cox and RSF models. This confirms that the functional trajectory adds significant discriminative power beyond scalar covariates.

These results show that incorporating the full performance trajectories as functional covariates leads to a clear improvement in ranking ability, beyond what can be achieved with scalar summaries of career length and observation density alone. At the same time, all models capture non-trivial structure.

4.2 Time-to-peak prediction and calibration

While hazard-based metrics focus on relative ordering, practical applications often require explicit predictions of the time-to-peak for individual athletes. As described in Section 3.6, we obtain calibrated time-to-peak predictions $\widehat{T}_i$ by fitting, within each event family and model, a smooth mapping from the training-set risk scores r_i to the observed times-to-peak T_i among uncensored athletes, and then applying this mapping to the test-set risk scores. Prediction accuracy is quantified using the root mean squared error (RMSE) and mean absolute error (MAE) among uncensored test-set athletes.

The bottom panels of Figure 3 and Table 1 report the calibrated RMSE and MAE. After

calibration, the AFCM achieves the lowest errors on average, indicating that nonlinear, phase-varying functional effects improve not only ranking performance but also absolute time prediction. The FLCM follows closely. The scalar Cox model and RSF generally exhibit higher prediction errors, confirming that the full longitudinal signal adds value for precise timing prediction.

The absolute magnitude of prediction errors—ranging from approximately 1.7 to 2.2 years in MAE and 2.3 to 2.8 years in RMSE across event families—reflects the substantial inherent variability in athletic career trajectories. Peak timing is influenced by a complex interplay of factors not captured in our data, including training regimes, injury history, competition schedules, and individual life circumstances. Nevertheless, all four model classes substantially outperform a naive baseline that always predicts the sample mean time-to-peak (baseline RMSE: 3.46 years): the AFCM achieves 31.9% error reduction, FLCM 29.8%, scalar Cox 24.5%, and RSF 15.1%. Within this context, the models' ability to consistently improve upon naive predictors and to achieve strong discrimination metrics demonstrates that performance trajectories carry meaningful signal about peak timing, even if point predictions remain uncertain. For practical applications such as talent identification and resource allocation, the relative ranking of athletes (captured by the C-index) may be more valuable than precise year-level forecasts.

Figure 4 provides a complementary view of calibration by plotting predicted versus observed time-to-peak for uncensored athletes, faceted by event family and model. After calibration, the scatterplots align reasonably well with the diagonal reference line, but residual patterns differ systematically between models. The functional models produce tight clouds around the diagonal, particularly in sprint, hurdles and middle-distance events, reflecting their ability to exploit both scalar covariates and trajectory shape. Scalar Cox and RSF predictions show more regression to the mean and larger dispersion in late-peaking athletes, consistent with their higher RMSE and MAE.

4.3 Summary of comparative performance

Taken together, the evaluation results reveal a coherent picture. Functional Cox models (FLCM and AFCM) dominate in hazard-based discrimination and also exhibit the best calibrated time-to-event predictions in most event families. This indicates that the shape of an athlete's performance trajectory carries substantial information about both *relative* and *absolute* peak timing beyond

17

what is captured by overall duration or observation density alone. At the same time, the scalar Cox model remains a strong baseline: with only two covariates and a simple calibration step it achieves performance close to that of the functional models, especially in families where trajectories are more homogeneous.
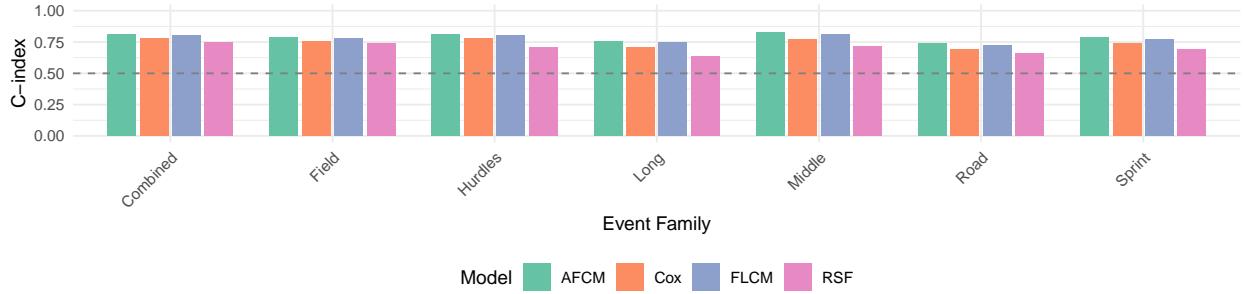
In practice, this suggests a trade-off. Scalar models may suffice when only coarse-grained predictions of peak timing are required or when computational simplicity is paramount. Functional models become attractive when ranking quality matters, when calibrated predictions are important, or when phase-specific interpretation of the trajectory is a scientific objective. We return to these points in Section 6.

Table 1: Average calibrated prediction error (RMSE and MAE) for each model class, averaged across all event families. The improvement percentage shows the reduction in RMSE relative to a naive baseline that always predicts the sample mean time-to-peak (3.46 years). All four model classes substantially outperform the baseline, with the AFCM achieving the largest error reduction (31.9%).

Model	RMSE (years)	MAE (years)	Improvement (%)
Baseline (Mean)	3.464	2.539	—
Scalar Cox	2.616	1.915	24.5
Random Forest	2.940	2.154	15.1
FLCM	2.432	1.782	29.8
AFCM	2.360	1.729	31.9
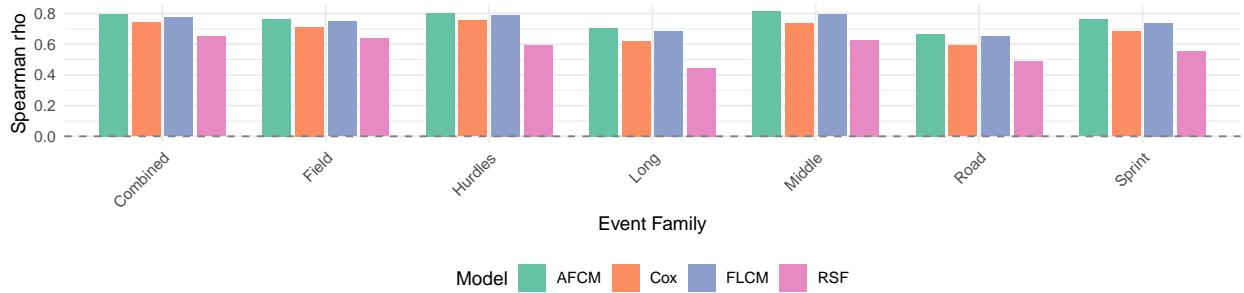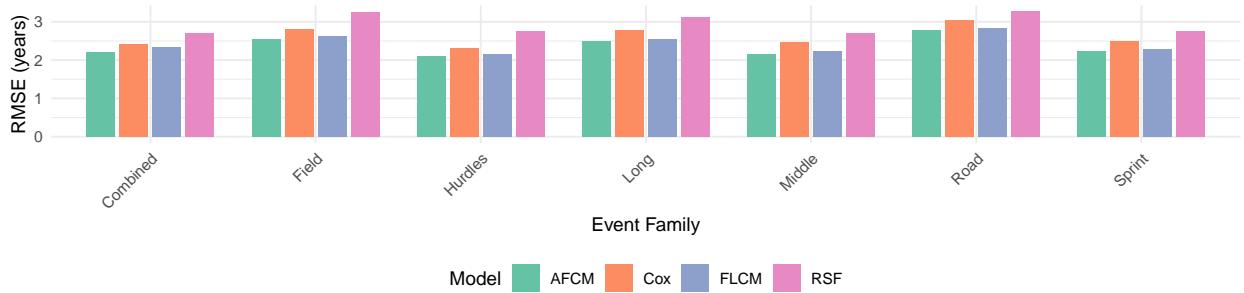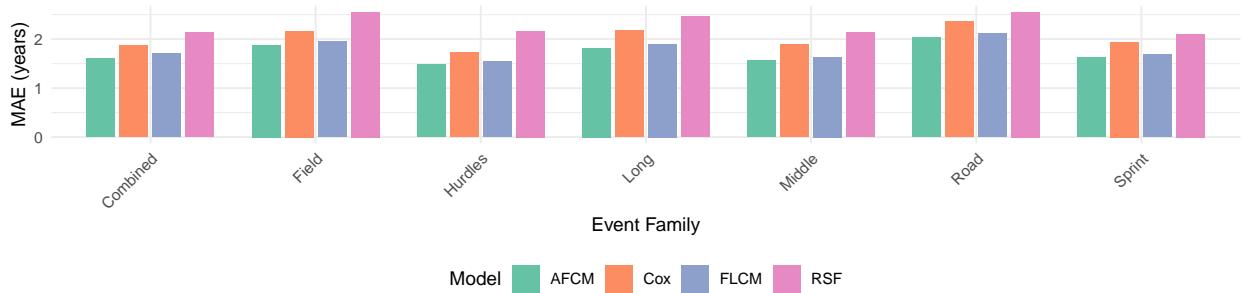
Figure 3: Comprehensive model evaluation across event families. The four panels display: (Top-Left) C-index for hazard-based discrimination; (Top-Right) Spearman rank correlation between risk scores and observed time-to-peak; (Bottom-Left) Root Mean Squared Error (RMSE) of calibrated time-to-event predictions; and (Bottom-Right) Mean Absolute Error (MAE) of calibrated predictions. Functional models (AFCM, FLCM) consistently outperform scalar baselines in ranking metrics and achieve the lowest prediction errors.
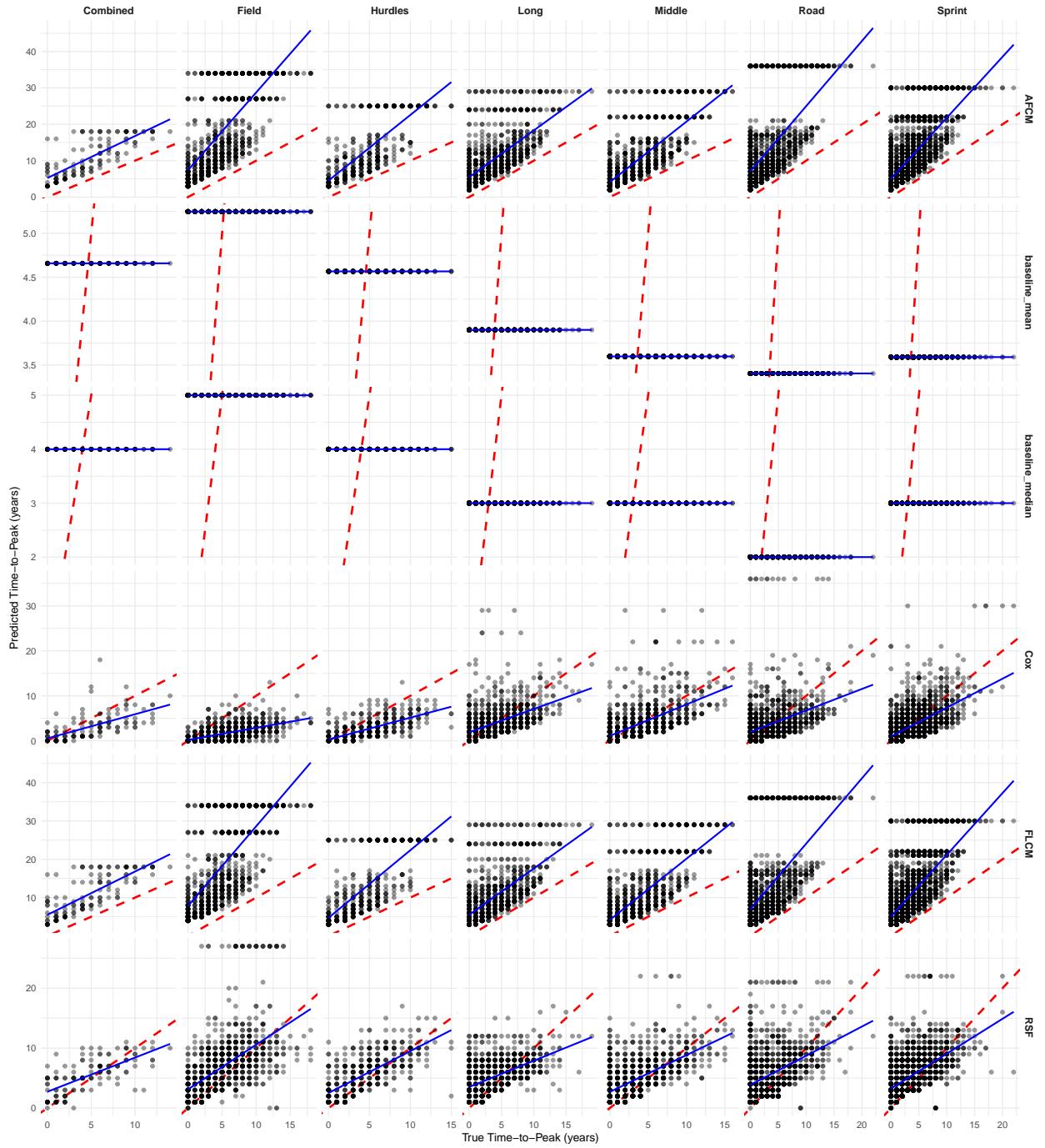
Figure 4: Predicted versus observed time-to-peak for uncensored athletes in the test set, stratified by event family (columns) and model (rows). The red dashed line indicates perfect calibration ($\widehat{T}_i = T_i$), and the blue line shows a fitted linear trend. Functional models provide good relative ordering and well-calibrated predictions, with tighter scatter around the diagonal than scalar Cox and RSF in most families.

5 Phase-specific functional effects

For brevity, we show the coefficient function for the Combined family; coefficient functions for the remaining families are provided in Appendix. To summarize the phase-specific contributions of the functional covariate, we discretize career phase into three equal intervals: Early ($s \in [0, 1/3)$), Mid ($s \in [1/3, 2/3)$), and Late ($s \in [2/3, 1]$). This tripartite division provides an interpretable summary of how the predictive importance of performance varies from career beginning to end. We visualize the estimated coefficient function $\hat{\beta}(s)$ in Figure 5, while Figure 6 provides a compact summary of the mean absolute effects by phase.

FLCM Coefficient Function: Combined
beta(s): Effect of performance at career phase s on hazard (with 95% CI)
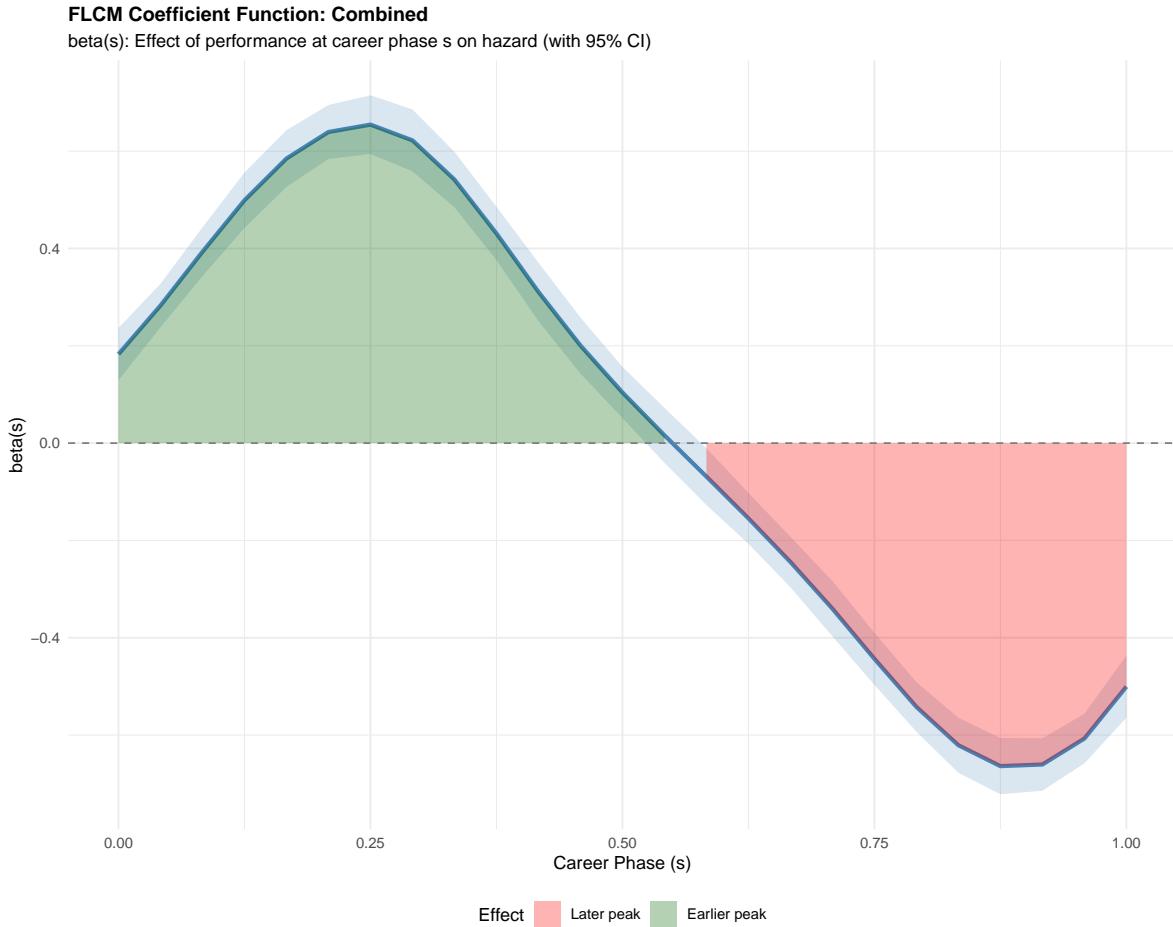
Figure 5: Estimated coefficient function $\hat{\beta}(s)$ (with 95% CI) from the FLCM for the combined family. Positive values imply that higher performance at that career phase is associated with earlier peaks (higher hazard), while negative values imply later peaks.
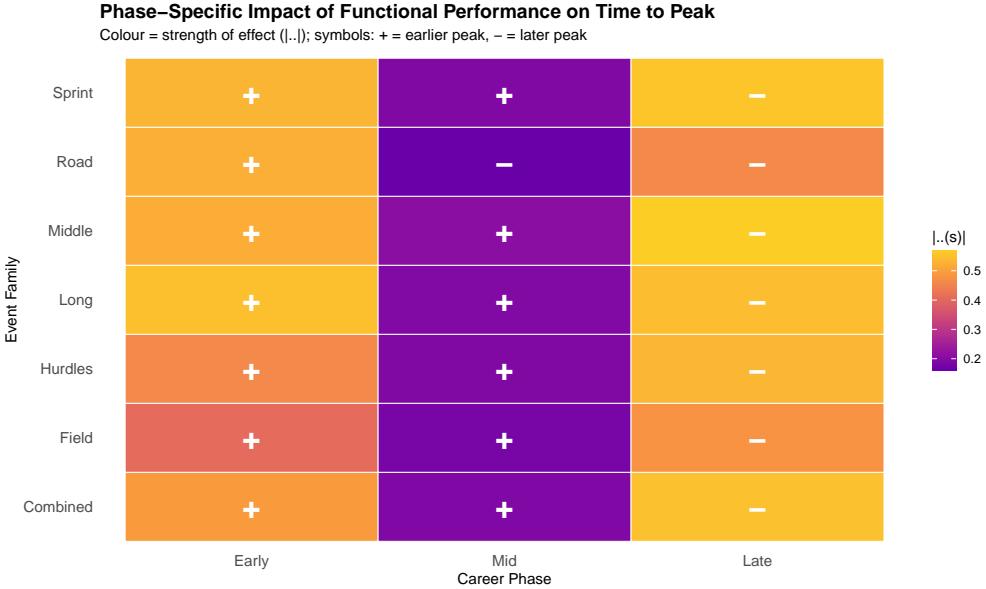
Figure 6: Phase-specific impact of functional performance on time-to-peak. Rows correspond to event families, columns to career phases (early, mid, late). Colour encodes the strength of the functional effect via $|\hat{\beta}(s)|$, with brighter tiles indicating phases where the trajectory is most informative. The sign indicates the direction of the effect: $+$ means that higher standardized performance during that phase is associated with an earlier peak (higher hazard), while $-$ means that higher performance is associated with a later peak (lower hazard).

5.1 Phase-specific functional insights

Bright tiles in Figure 6 (large $|\hat{\beta}|$) identify career phases in which the within-career performance trajectory carries strong signal about when an athlete will peak. These are precisely the regions where functional models extract information that scalar summaries cannot capture.

The sign encodes the direction of the effect. Positive signs $(+)$ indicate phases where $\hat{\beta}(s) > 0$: athletes who perform above their own typical level in these phases tend to reach their peak earlier, corresponding to an increased hazard of peaking. Negative signs $(-)$ indicate phases where $\hat{\beta}(s) < 0$: here, above-typical performance is associated with delayed peaks and a reduced hazard. Across most event families, early-career performance has a positive association with earlier peaks, whereas late-career performance tends to have a protective (negative) effect, consistent with the idea that sustained high performance later in a career is characteristic of athletes whose peak occurs later.

5.2 Scientific value of functional modelling

A purely scalar Cox model based on career length and log number of observations cannot distinguish *which* parts of the trajectory matter for peak timing: it can only summarise the overall level or duration of performance. The FLCM, by contrast, decomposes the predictive contribution of the trajectory over career phase and reveals phase-specific dynamics that are invisible to scalar models. Even in settings where overall predictive performance metrics (e.g. C-index, RMSE) for functional and scalar models are similar, the functional model delivers interpretable, phase-resolved information about *when* in an individual's trajectory performance is most predictive of event time. This phase-specific insight is the main scientific gain from functional modelling in our application and is directly transferable to other domains where understanding *when* along a longitudinal process risk accumulates or dissipates is of interest.

6 Discussion

We have examined functional Cox regression as a tool for time-to-event prediction from longitudinal trajectories, using elite track-and-field performance as a case study. The results highlight both the potential and the practical considerations of functional survival models.

Incorporating full performance trajectories as functional covariates yields consistent gains over strong scalar baselines. Functional Cox models achieve the best discrimination and, after calibration, the lowest time-to-peak prediction errors in most event families. This indicates that the *shape* of a trajectory contains genuine predictive information about event timing beyond what can be captured by overall duration or observation density alone. The phase-specific coefficient functions reveal interpretable patterns: early-career performance typically predicts earlier peaks, while sustained high performance later in a career is associated with delayed peaks. These insights are invisible to scalar models and represent the main scientific value of the functional approach.

Our analysis has several limitations. First, we focus exclusively on season-best performances, discarding within-season variation that may contain additional signal. Second, the data lack contextual information such as training load, injury history, or coaching changes that undoubtedly influence peak timing. Third, our censoring mechanism treats all post-2023 careers as censored, which may underestimate peak timing for very recent entrants. Future work could explore joint

models that incorporate time-varying covariates, extend the framework to recurrent events (e.g. multiple peaks), or investigate whether similar phase-specific effects emerge in domains such as clinical biomarker trajectories or equipment degradation curves.

While our case study focuses on athletic performance, the methodological framework generalizes naturally to other settings with irregular longitudinal monitoring and time-to-event outcomes. In medicine, one might model the trajectory of a biomarker (e.g. glucose, blood pressure) to predict time to disease onset, revealing which phases of the patient's monitoring history are most informative. In reliability engineering, sensor readings over equipment lifetime could predict time to failure, identifying critical degradation phases. In digital platforms, user engagement trajectories could predict time to churn. In each domain, the functional Cox framework offers both improved prediction and interpretable, phase-resolved insights into *when* along the trajectory risk accumulates.

7   Conclusion

This paper demonstrated that functional Cox regression provides a powerful and interpretable approach for time-to-event prediction from irregular longitudinal trajectories. By treating each trajectory as a functional covariate, we achieve superior discrimination and calibrated predictions compared to scalar Cox and random forest baselines, while simultaneously uncovering phase-specific effects that reveal *which* parts of the trajectory are most predictive of event timing.

In our track-and-field case study, these functional effects align with substantive notions of "early peakers" versus "late developers", and the estimated coefficient functions provide actionable insights for talent identification and career planning. More broadly, the framework is directly applicable to any domain where understanding the temporal structure of longitudinal risk is scientifically or practically valuable.

We release our full R code and aggregated data to facilitate reproducibility and extensions. Functional survival analysis stands as a natural bridge between complex longitudinal data and interpretable time-to-event modelling, with broad applicability across sport science, medicine, reliability engineering, and behavioural analytics.

A   Appendix: Additional Coefficient Functions

Figure 7 displays the estimated coefficient functions for the remaining event families.
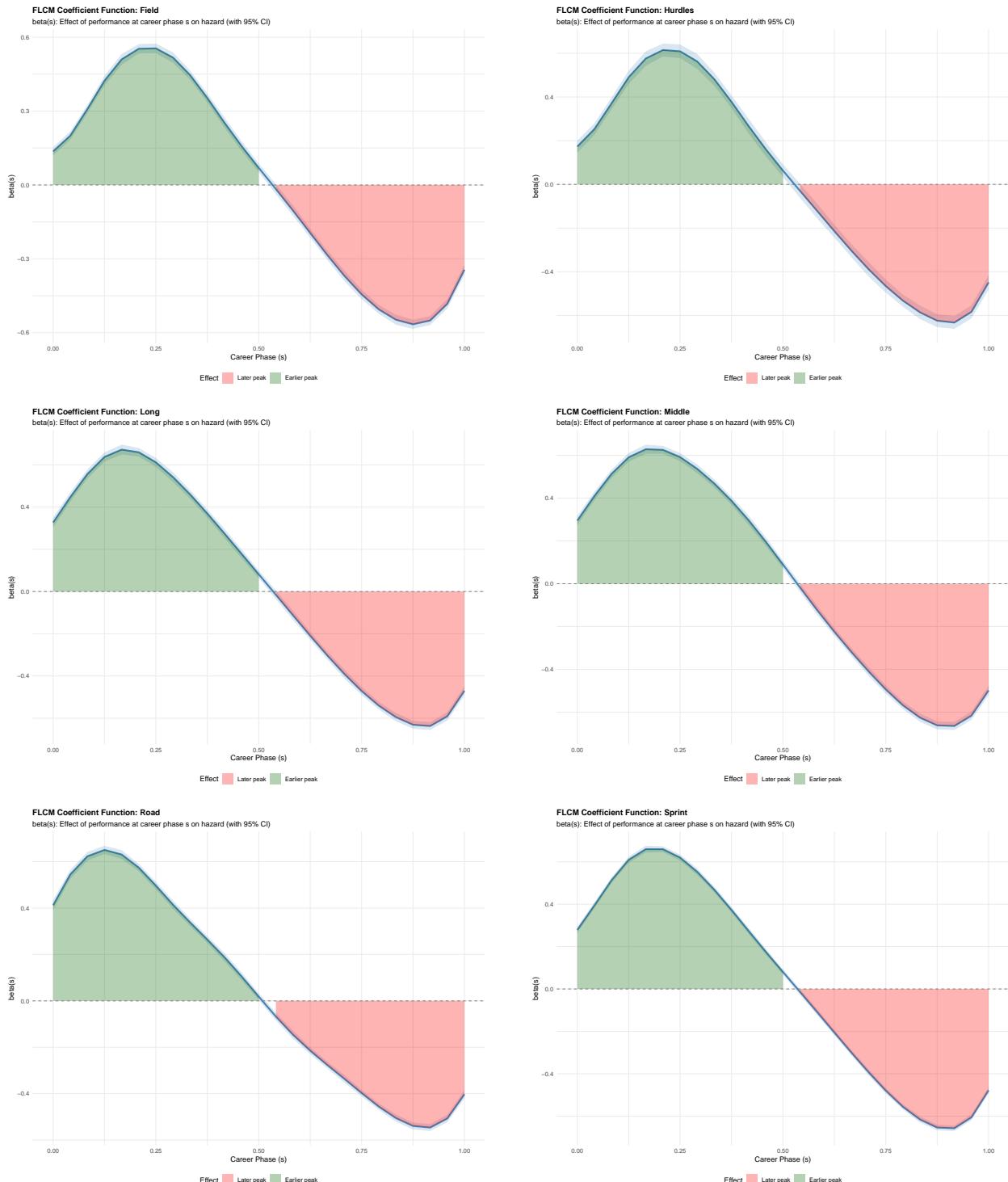
Figure 7: Estimated FLCM coefficient functions for the remaining event families.

References

David R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society (Series B)*, 34(2):187–220, 1972.

Frank E. Harrell, Robert M. Califf, David B. Pryor, Kerry L. Lee, and Robert A. Rosati. Evaluating the yield of medical tests. *JAMA*, 247(18):2543–2546, 1982.

Hemant Ishwaran, Udaya B. Kogalur, Eugene H. Blackstone, and Michael S. Lauer. Random survival forests. *The Annals of Applied Statistics*, 2(3):841–860, 2008.

Andrada E. Ivanescu, Ana-Maria Staicu, Fabian Scheipl, and Sonja Greven. Penalized function-on-function regression. *Computational Statistics*, 30:539–568, 2015.

D. Kong, J. G. Ibrahim, E. Lee, and H. Zhu. Functional linear cox regression model. *Biometrics*, 74(1):109–117, 2018.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL `https://www.R-project.org/`.

James O. Ramsay and Bernard W. Silverman. *Functional Data Analysis*. Springer, 2005.

Fabian Scheipl, Ana-Maria Staicu, and Sonja Greven. Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2):477–501, 2015.

Hajime Uno, Tianxi Cai, Michael J. Pencina, Ralph B. D'Agostino, and L. J. Wei. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, 30(10):1105–1117, 2011.

Simon N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (Series B)*, 73 (1):3–36, 2011.