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, usage patterns or sensor readings. Standard survival models typically compress these trajectories into a few scalar summaries, sacrificing information about *which* phases of the trajectory are most predictive for event timing. Black-box machine learning methods can exploit rich features but rarely provide phase-specific interpretation.

We develop and empirically study a functional Cox regression framework for time-to-event prediction that treats each individual 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 of the trajectory contribute to the hazard. We combine this with scalar covariates and compare four model classes: a scalar Cox proportional hazards model, a random survival forest, a functional linear Cox model (FLCM) and an additive functional Cox model (AFCM) with nonlinear functional effects.

As a case study we analyse more than 1.3 million season-best performances from 67,977 athlete–event careers in elite track-and-field, where the goal is to predict the time from event debut to peak performance. Models are evaluated within broad event families using both hazard-based discrimination metrics and calibrated time-to-peak predictions. Across families, functional Cox models consistently achieve the best or near-best concordance and rank correlation, showing that trajectory shape contains substantial information beyond simple summaries of career length and observation density. After a one-dimensional calibration step mapping risk scores to time, functional models also attain the lowest root mean squared error and mean absolute error in most families, although scalar Cox remains competitive.

The estimated functional coefficient $\beta(s)$ can be summarised over interpretable early, mid and late phases, highlighting career phases where above-typical performance is associated with earlier or later peaks. These phase-specific effects represent information that cannot be extracted from scalar models alone. 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.

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 [**?**]. 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.

In this paper we develop 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 more than one million season-best results, we construct standardized annual 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) We show that functional Cox models can yield consistent improvements in discrimination and calibrated prediction over strong scalar baselines, while also delivering phase-specific coefficient functions that identify which parts of a trajectory are most informative for the event time.

(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 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, consisting of annual season-best results for male and female athletes across 93 distinct events. After preprocessing, the working dataset contains $N_{\mathrm{AE}} = 67{,}977$ athlete–event combinations and more than 1.3 million yearly observations.

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.
- **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 construct a standardized performance metric that maps each annual season-best result to a dimensionless score on which larger values consistently correspond to better performance. Concretely, for each event–sex–year stratum we compute a location-scale transformation of raw results based on the empirical distribution of all athletes in that stratum. We then reverse signs as needed so that higher values always indicate better performance. The resulting standardized score can be interpreted as a performance level relative to contemporaneous peers in the same event.

For each athlete–event combination, we extract the standardized annual season-best score for each calendar 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.

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

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

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

For each athlete–event i, we identify a *career 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. We then define the time-to-peak in years as

$$T_i = j_i^\star - 1,$$

measured from the start of the career in that event. If the last observed year occurs in or after the final calendar year of the database (2023) and the peak is attained in the final observed year, the athlete may still be competing and the peak timing beyond the observation window is unknown.

We therefore treat such cases as right-censored:

$$\Delta_i = \begin{cases} 1, & \text{if the peak occurs before the end of follow-up,} \\ 0, & \text{if the athlete is still active at the end of follow-up.} \end{cases}$$

For censored athletes we set T_i equal to the observed time from first to last appearance in the event. Formal inclusion criteria exclude pathological cases where the implied time-to-peak would be negative or the trajectory does not exhibit a meaningful rise.

Across all event families, approximately 80% of athlete–event combinations have an observed peak within the follow-up window ($\Delta_i = 1$), while the remaining 20% are right censored due to ongoing careers. The mean time-to-peak is around 3–4 years from event debut, with variation across event families.

Figure 1 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.

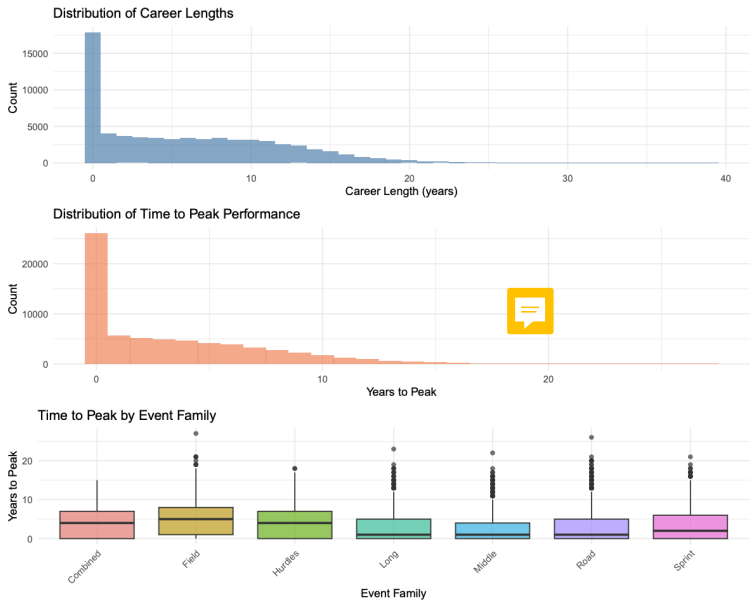

Figure 1: Distributions of career length and time to peak performance. Top: histogram of career length (years) across all athlete–event combinations. Middle: histogram of time from first appearance in the event to peak performance. Bottom: boxplots of time to peak stratified by event family.

Figure 2 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 not yet peaking, whereas road and long-distance athletes maintain a higher pre-peak probability at later times. A global log-rank test strongly rejects equality of the survivor functions ($p < 0.0001$).

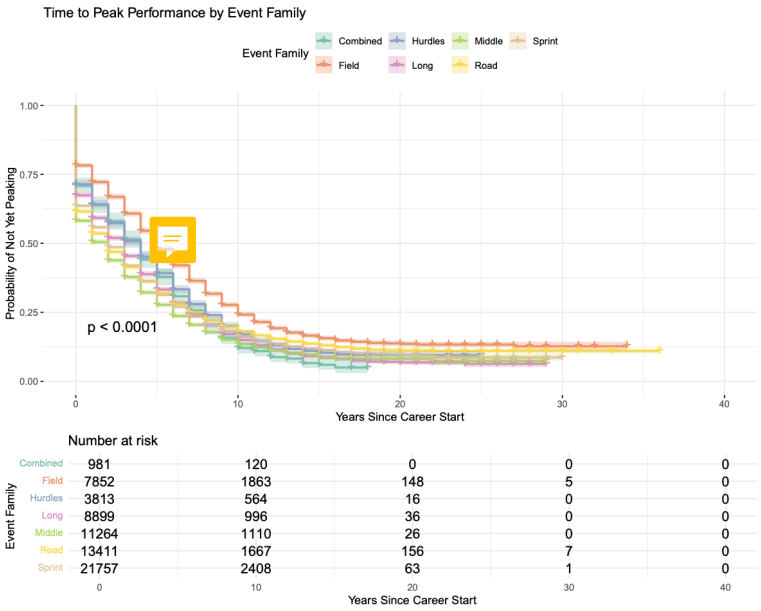

Figure 2: Kaplan–Meier estimates of the survivor function for time to peak performance, stratified by event family. The curves give the probability that an athlete has not yet reached their observed peak as a function of years since entering the event. A global log-rank test ($p < 0.0001$) rejects equality of curves, indicating substantial heterogeneity in peak timing across families. The risk table shows the number of active careers remaining at selected time points.

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. 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 3 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.

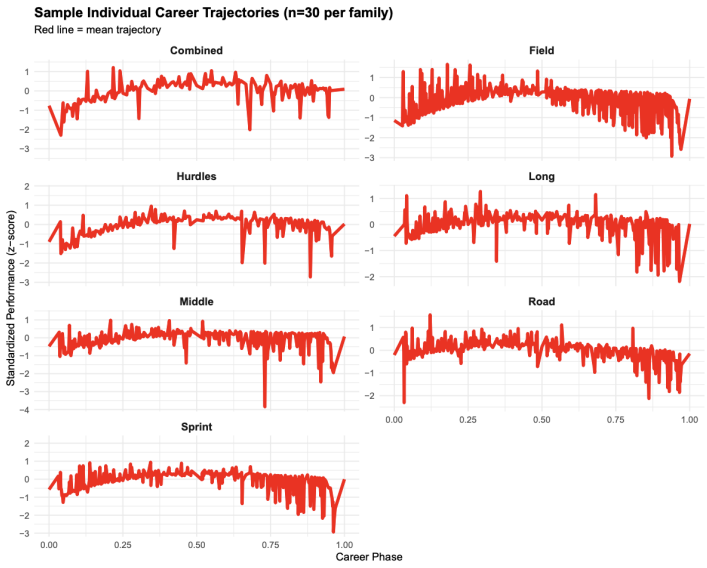

Figure 3: Sample standardized performance trajectories by event family. For each family, 30 athlete–event careers are displayed over normalized career phase ($s \in [0, 1]$); the red line denotes the mean trajectory. Typical patterns involve early improvement, a mid-career plateau, and eventual decline, with systematic differences in shape across families.

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 overall performance level and career length.

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 separately within each event family. Within families, PH diagnostics 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 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 (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* r_i^{RSF}, for example by taking the integrated cumulative hazard over a fixed time window or the expected event time. 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. 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 a roughness penalty on $\beta(s)$, for example $\int_0^1 \{\beta''(s)\}^2 ds$. This yields a smooth estimate $\widehat{\beta}(s)$ that 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:

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

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 Training and cross-validation

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 yields a scalar risk score r_i for each athlete–event i in the test set, where larger values of r_i correspond to an increased hazard of reaching the peak earlier. For Cox models, $r_i = \widehat{\gamma}^\top Z_i$ for the scalar Cox and $r_i = \widehat{\gamma}^\top Z_i + \int_0^1 \widehat{\beta}(s) X_i(s) ds$ for the FLCM. For the RSF and AFCM we similarly extract monotone risk scores from the fitted models.

We assess discrimination using Harrell's concordance index C, defined as the probability that, for a randomly chosen pair of athletes with ordered event times, the athlete who peaks earlier has a higher predicted risk. For censored data we use the usual comparable-pairs definition. We also compute Spearman's rank correlation between the risk scores $\{r_i\}$ and observed times-to-peak $\{T_i\}$ among uncensored 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 an explicit prediction 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 treat the model's risk score r_i on the training set as a one-dimensional predictor of the observed time-to-peak T_i among uncensored athletes. We then fit a flexible generalized additive model (GAM)

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

where $g(\cdot)$ is a smooth function estimated via penalized splines and ε_i is an error term. The fitted function $\widehat{g}$ maps risk scores to calibrated time-to-peak predictions. On the test set, we obtain predicted times $\widehat{T}_i = \widehat{g}(r_i)$.

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

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

and the mean absolute error (MAE)

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

where N_{unc} is the number of uncensored athletes in the test set for the given family. This separation between discrimination and calibration allows us to distinguish models that rank athletes well from those that provide reliable absolute time predictions.

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 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 4 (top two panels) summarizes the hazard-based metrics across event families and models. In every family the FLCM attains the highest or near-highest C-index, typically in the range 0.74–0.86, indicating substantially better-than-random ordering of athletes by time-to-peak. The AFCM achieves very similar discrimination, whereas the scalar Cox model is consistently a little worse and RSF is uniformly the weakest performer, often with C-indices below 0.70. Spearman correlations tell the same story: functional models yield the largest (in absolute value) correlations between risk scores and time-to-peak, while RSF again trails behind.

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, and there is no indication of pathological fits that would suggest coding errors.

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 monotone 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 two panels of Figure 4 display the calibrated RMSE and MAE across families and models. Calibration reduces errors substantially for all models compared with the raw survival-based predictions, bringing RMSE and MAE into a plausible range of roughly 2–4 years depending on family. After calibration, functional Cox models generally achieve the lowest errors: in most families the FLCM has the smallest RMSE and MAE, with the AFCM performing comparably, while the scalar Cox model and RSF exhibit systematically larger prediction errors. The gaps are not enormous—on the order of a few tenths of a year in MAE for many families—but they are consistent in direction.

Figure 5 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 what is captured by career length and 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.

Figure 4: Summary of test-set performance across event families and models. Top: C-index for hazard-based discrimination (higher is better; horizontal dashed line at 0.5 corresponds to random ranking). Second panel: Spearman rank correlation between risk scores and observed time-to-peak among uncensored athletes (larger magnitude is better). Bottom panels: RMSE and MAE for calibrated time-to-peak predictions among uncensored test-set athletes.

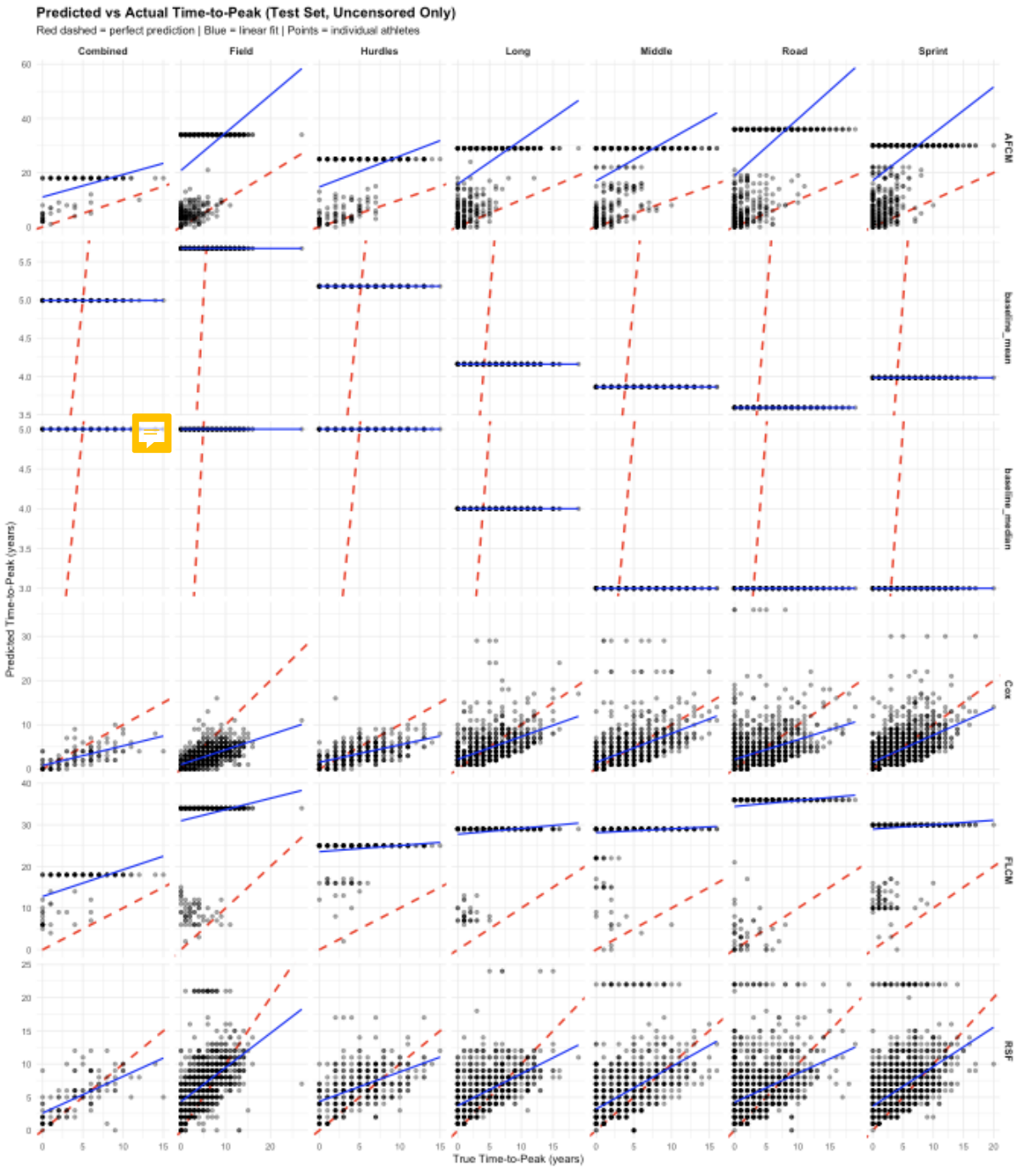

Figure 5: Predicted versus observed time-to-peak for uncensored athletes in the test set, stratified by event family (columns) and model (rows). The red dashed line indicates perfect calibration ($\widehat{T}_i = T_i$), and the blue line shows a fitted linear trend. Functional models provide good relative ordering and well-calibrated predictions, with tighter scatter around the diagonal than scalar Cox and RSF in most families.

5 Phase-specific functional effects

To summarize the phase-specific contributions of the functional covariate, we discretize career phase into three intervals (early, mid, late) and average the estimated coefficient function $\hat{\beta}(s)$ of the FLCM within each phase and event family. Figure 6 reports the mean coefficient mean$\{\hat{\beta}(s)\}$ and its mean absolute value mean$\{|\hat{\beta}(s)|\}$ by phase, while Figure 7 provides a compact visualization in which colour encodes $|\hat{\beta}(s)|$ and arrows indicate the sign of $\hat{\beta}(s)$.

event_family	phase	mean_beta	mean_abs_beta
<chr>	<fctr>	<dbl>	<dbl>
Combined	Early	0.355544701	0.3555447
Combined	Mid	0.036770067	0.1218881
Combined	Late	-0.440560101	0.4405601
Field	Early	0.275797172	0.2757972
Field	Mid	0.025564640	0.1925921
Field	Late	-0.430996025	0.4309960
Hurdles	Early	0.283769464	0.2837695
Hurdles	Mid	0.036906582	0.1858135
Hurdles	Late	-0.443920074	0.4439201
Long	Early	0.397026325	0.3970263
Long	Mid	-0.015970176	0.1542185
Long	Late	-0.485581571	0.4855816
Middle	Early	0.415454312	0.4154543
Middle	Mid	0.016662542	0.1857163
Middle	Late	-0.496301761	0.4963018
Road	Early	0.466377006	0.4663770
Road	Mid	-0.028762589	0.1374637
Road	Late	-0.421878305	0.4218783
Sprint	Early	0.368482313	0.3684823
Sprint	Mid	-0.001344844	0.1620817
Sprint	Late	-0.451596785	0.4515968

Figure 6: Phase-averaged functional effect estimates from the FLCM. For each event family and career phase (early, mid, late), the table reports the mean coefficient mean$\{\hat{\beta}(s)\}$ and the mean absolute coefficient mean$\{|\hat{\beta}(s)|\}$ over that phase. Larger absolute values indicate phases in which the functional trajectory carries stronger information about time-to-peak.

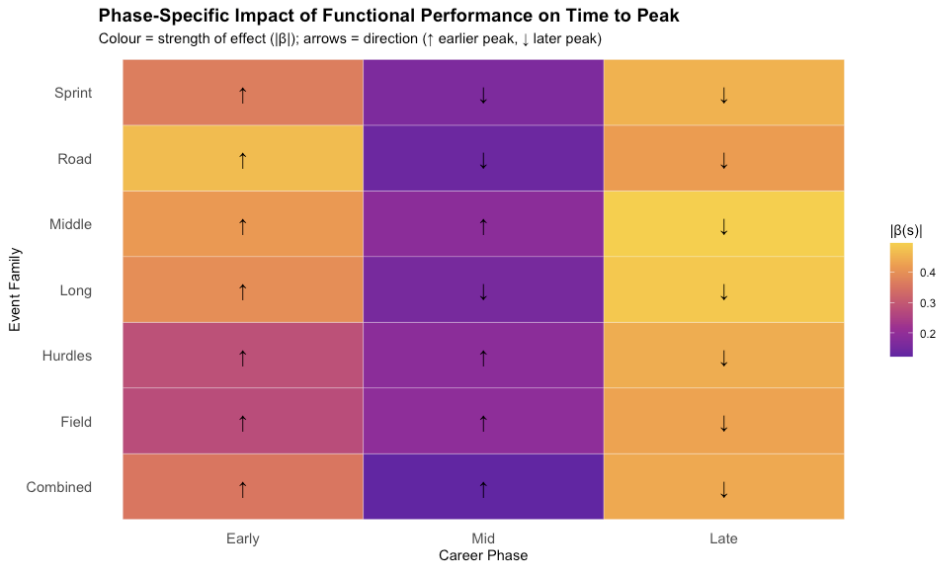

Figure 7: Phase-specific impact of functional performance on time-to-peak. Rows correspond to event families, columns to career phases (early, mid, late). Colour encodes the strength of the functional effect via $|\hat{\beta}(s)|$, with brighter tiles indicating phases where the trajectory is most informative. Arrows indicate 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 7 (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 arrows encode the direction of the effect. Upward arrows 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. Downward arrows 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 limitations of functional survival models.

On the positive side, 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.

Perhaps more importantly, the functional models provide structured, phase-specific insight through the estimated coefficient function $\beta(s)$. By summarising $\beta(s)$ over early, mid and late phases we can quantify when along the trajectory above-typical performance is associated with earlier or later peaks. In our case study, early high performance tends to accelerate peak timing, while sustained high performance in later phases is associated with delayed peaks. This formalises intuitive notions of "early peakers" and "late developers" in a way that scalar models cannot, and suggests concrete hypotheses for training and talent-development strategies.

At the same time, the gains in classical prediction metrics are modest rather than dramatic. After calibration, RMSE and MAE for functional and scalar Cox models differ by at most a few tenths of a year in many families. This reflects both the strength of the scalar covariates (which already carry substantial information about peak timing) and the intrinsic difficulty of predicting precise event times from noisy longitudinal data. In applications where only coarse-grained timing is needed, a well-specified scalar Cox model may therefore remain an attractive baseline.

Our study also underscores several practical challenges. First, the quality of functional modelling depends critically on preprocessing choices: time normalisation, smoothing, within-individual standardisation and basis dimension all affect the estimated coefficient functions. Second, survival curves from flexible functional models can be poorly calibrated in absolute time, motivating the use of simple post-hoc calibration steps such as the GAM mapping adopted here. Third, our analysis uses a single functional predictor and low-dimensional scalar covariates; many real-world applications involve multiple correlated streams (e.g., several biomarkers or sensor channels) and higher-dimensional scalar information.

Looking beyond sport, the methodological framework is directly applicable to other domains where rich longitudinal signals precede an event. In medicine, trajectories of lab tests or vital signs could be modelled functionally to study which phases of disease progression are most predictive of hospitalisation or mortality. In reliability, sensor-based degradation curves could be linked to failure times. In digital platforms, engagement histories could be used to understand which stages

of user behaviour are most predictive of churn. In all of these examples functional Cox models can be used not only to improve predictions, but also to answer scientifically meaningful questions about *when* along a process risk is concentrated.

Future work could extend our approach in several directions. One obvious step is to handle multiple functional covariates jointly, using multi-dimensional functional bases or low-rank tensor decompositions. Another is to embed functional Cox models in hierarchical or Bayesian frameworks that borrow strength across related groups and propagate uncertainty more fully. Finally, integrating richer non-survival information (such as training load, injury history or physiological measurements in the sports setting) would allow a more complete assessment of how longitudinal trajectories interact with other factors to determine event timing.

7 Conclusion

We have proposed and empirically evaluated a functional survival modelling framework for time-to-event prediction from longitudinal trajectories, with a large-scale case study on time-to-peak performance in elite track-and-field. By framing time-to-peak as a right-censored outcome and combining scalar and functional covariates, we provided a unified comparison of scalar Cox, random survival forest, functional linear Cox and additive functional Cox models across multiple event families.

Our results show that functional Cox models can leverage the full shape of longitudinal trajectories to achieve strong discrimination and competitive, well-calibrated time-to-event predictions, while also delivering phase-specific coefficient functions that identify when along the trajectory performance matters most. Scalar Cox models remain strong baselines, particularly when only simple covariates are available, but they cannot recover these phase-specific effects.

The methodology is not specific to sport. Any application that combines a time-to-event outcome with longitudinal measurements observed on a comparable time scale can, in principle, benefit from functional Cox regression and the accompanying calibration and summarisation tools described here. As longitudinal data become increasingly pervasive in medicine, industry and the social sciences, we expect functional time-to-event models to play an important role in building predictive systems that are both accurate and interpretable.

References