MDPI

Article

Modeling Workload and Injury Risk in Elite Touch Rugby with Clustering Effect: A Time-Scaled Shared Frailty Approach

Tom Huang [1], Shu Su [2], Nuttanan Wichitaksorn [3]* and Kirsten Spencer [4]

[1] Department of Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand; thsy8599@gmail.com
[2] Department of Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand; shu.su@aut.ac.nz
[3] Department of Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand; nuttanan.wichitaksorn@aut.ac.nz
[4] School of Sport, Exercise & Health, Auckland University of Technology, Auckland, New Zealand; kirsten.spencer@aut.ac.nz
* Correspondence: nuttanan.wichitaksorn@aut.ac.nz

Abstract

In this study, we propose a general mathematical modeling framework based on the characteristics of elite athletes' movements in the touch ruby matches to investigate the dynamics relationship between physical workload and injury risk over time. Our framework extends the frailty model by incorporating a time-scaling component, allowing for a personalized baseline hazard while capturing the dynamic effects of covariates (e.g., velocity variation) on athletes' survival times and cluster-specific injury vulnerability. We applied our model to high-frequency wearable sensor data collected from 27 elite athletes (15 men and 12 women). The empirical study results show that our model, time scaled frailty mode (TSFM), demonstrate superior goodness-of-fit than traditional frailty and Andersen–Gill models. The results reveal that higher velocity variation, particularly during high-intensity phases, and longer time of continuous exposure to the workload spike state significantly increased overload risk, ultimately resulting in injury. It also highlights the importance of individual differences, even under the same exercise intensity. These insights provide coaches with an evidence-based framework for athlete monitoring, allowing for more personalized training loads, tactical deployment, and injury prevention strategies in elite Touch Rugby environments.

Keywords: survival analysis; frailty model; recurrent events modeling; unobserved heterogeneity; external workload analysis; piecewise exponential additive mixed model

1. Introduction

Touch rugby is a fast-paced and intermittent team sport that combines high intensity movement with frequent transitions between offense and defense. Unlike traditional rugby, strong physical tackle is replaced by light hand touch, focusing on agility and sprinting. Matches are typically played on a 70 m × 50 m field with six athletes per team, which allows unlimited substitutions throughout the match that are two 20-minute halves [1]. Despite the removal of strong physical tackles in touch rugby, the sport still imposes considerable physical demands, including intermittent high-intensity efforts and rapid accelerations. Analyzing physical demand is essential, which offers deeper insights into the workload and performance of every athlete on the team. To accurately capture and monitor these demands, wearable technology has become an pivotal solution.

Received:
Revised:
Accepted:
Published:

Recent research has shown that wearable technology plays an increasingly important role in monitoring and tracking athletes' movements, as well as collecting these data, during training and matches, which are subsequently used to improve training and optimize match strategies effectively [2–4]. The advancement of wearable tracking systems, including GPS and inertial sensors, can enable more comprehensive analysis of athletes' workloads. Therefore, these systems have become widely adopted in the field of sport science to help mitigate injury risk and optimize match strategies [5]. In particular, the metrics obtained from these wearable tracking systems, including distance, velocity, and other related measurements, can be used to estimate the internal and external workloads of athletes [6–9]. So, it can help coaches understand how hard athletes work, adjust training plans, and set up a reasonable recovery plan. For example, Beaven et al. employ GPS data to quantify physical demand by capturing the velocity of elite male athletes, and identify how frequently athletes transition between different intensity levels [10]. In addition, devices with embedded GPS technology have been widely used to capture athlete movement profiles in various team sports, such as football, rugby, etc.[11,12]. These data have been used to measure athletes' activities and analyze their work rate patterns during matches and training, allowing coaches to better understand the physical condition of each member of the team and develop a data driven match strategy [11].

Although GPS technology has been widely used in court-base sports, its reliability has raised concerns and been discussed in the recent literature. Jennings et al. show that GPS with a higher frequent sampling rate is more reliable and precise in measuring distance [13]. In contrast, the GPS device with a frequency of 1-Hz has been verified to not accurately capture movement data, which takes less than 1 second [2]. The more recent GPS device with 10-Hz allows us to measure small changes in velocity changes with a good accuracy level [14,15]. The reliability of a 10-Hz GPS unit for monitoring quick acceleration and deceleration has been further validated by Nikolaidis et al. [16].

Undoubtedly, with the rapid development of technology, especially in wearable sensors (GPS), athlete performance analysis has entered a new era in touch rugby. The availability of GPS data with detailed information allows researchers and coaches to go beyond simple summaries, by using statistical and mathematical models. For example, Beaven et al., Chow, Vickery and Harkness employed descriptive statistics and basic statistical tests, such as t-tests and ANOVA, on GPS data, to assess differences between playing positions [10,17,18]. They found that wings generally covered greater distances while middles had higher intensity metrics. Moreover, Vickery and Harkness used GPS data collected from national touch tournaments to investigate mixed gender competitions, focusing on differences in workload intensity between genders, and provide a suggestion on adjustments in training programs [18]. Zaragoza et al. provided further analysis on gender-specific workload demands at the international level, emphasizing the essential nature of individualized training thresholds for male and female athletes [19].

Although these studies provided valuable insights on workload demand, they did not have sufficient analytical depth to the temporal dynamics of fatigue among athletes throughout the duration of the game and to assess the risk of injury in touch rugby [20]. Furthermore, in contrast to internal workload, external workload can be measured more consistently and practically [21]. Therefore, our study will analyze the external workload consumption pattern of New Zealand touch rugby teams, considering both male and female athletes, based on high frequency (10-Hz GPS) data and develop a comprehensive workload–injury modeling framework to access the relationship between external physical workload and the probability of injury based on survival analysis, which can help formulate more scientific training and competition strategies, and support the development of better injury prevention practices.

The remaining structure of the paper is organized as follows: Section 2 synthesizes the extant literature on workload–injury analysis and modeling approaches. Section 3 develops a novel and comprehensive framework for modeling workload–injury in touch-rugby. Section 4 describes and analyzes the tough rugby data collected from New Zealand matches. Section 5 shows the results of our models and discusses the key findings of the results. Section 6 concludes.

2. Related Work

Research examining the triadic relationship between load, fatigue, and injury risk demonstrates that external workload acts as a driver of fatigue accumulation and subsequent injury [22–25]. Several metrics, such as acute chronic workload ratio (ACWR), session Rating of Perceived Exertion (sRPE), distance, high-speed running, etc., have been developed to assess this relationship [11,24,26,27]. External workload metrics, such as high-speed running distance, etc., acquired via wearable devices, provide objective and sport-specific measures, which do not directly capture athlete fatigue levels, since baseline fitness levels vary significantly between athletes, potentially can confound the analytical results [28,29]. Unlike standard workload metrics, ACWR integrates recent and historical information, so it can offer an individualized measure of physical demands and refine injury risk assessment [24]. Consistent with the workload–injury nexus, Hulin et al. revealed that an ACWR value greater than 1.5 indicates a workload spike and is associated with a greater probability of injury occurrence [30].

Although these studies offered pivotal findings within the workload–injury nexus, they provided limited mechanistic insight into understanding the interplay between injury and the temporal dynamics of athlete fatigue levels. To address this problem, recent research in this area has shifted the focus to more sophisticated statistical frameworks, most notably survival analysis. This approach models the instantaneous hazard of injury through time-to-event analysis, serving as a powerful tool for risk assessment in sports science Nielsen et al.. They also emphasized that survival analysis techniques are well-suited to explore fundamental questions, including how workload influences injury risk over time, how multiple injuries accumulate, and how to deal with competing risks in injury data [31].

In the early years, researchers provided a non-parametric way to estimate survival curves from censored data [32]. In 1972, Mantel introduced the log-rank test and provided a formal statistical method to compare survival curves among groups, allowing researchers to test whether differences in time-to-injury trajectories, such as injury incidence patterns across distinct workload quartiles or athlete positions [33]. During the same year, a landmark model in the field of survival analysis is the proportional hazards model introduced by Cox. He proposed a hazard function, allowing covariates to have a multiplicative effect on the baseline hazard, that is,

$$\lambda(t;\mathbf{z}) = \lambda_0(t)e^{\mathbf{z}\boldsymbol{\beta}},$$

where $\mathbf{z} = (z_1,\ldots,z_p)$ denotes p covariate values; $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^\top$ is the vector of unknown parameters; $\lambda(t)$ is the hazard failure rate at failure time T,

$$\lambda(t) = \lim_{\Delta t \to 0+} \frac{\mathrm{pr}(t \le T < t + \Delta t)}{\Delta t};$$

and $\lambda_0(t)$ is the unknown hazard function that is defined under the condition $\mathbf{z} = \mathbf{0}$ [34].

The Cox model has quickly become the dominant tool for analyzing time-to-event outcomes due to their robustness, methodological flexibility, and interpretability. For example, Gabbe et al. used a Cox regression to identify risk factors by adopting a time-

to-event approach that centers on the time to the first hamstring strain, which is the accumulation of on-field exposure hours prior to the event, in community Australian football [35]. It demonstrated the imperative of incorporating a temporal dimension into the model in injury analysis. In addition, researchers utilized other approaches, such as the Kaplan–Meier estimator, the exponential survival model, Weibull survival model, etc., to study the time-to-first-event from different aspects, including survival probability and covariate effects [36–38]. These methods allow for the investigation of injury timing and athlete-specific risk, offering a more comprehensive approach to injury prediction and workload management [39]. The majority of survival models centered on modeling the duration of event-free participation prior to the first failure event; however, limited empirical attention has been directed toward the modeling of recurrent events in the sport science area [39,40]. Focusing only on the time to the first injury may not capture the complexity of injury risk well and the risk dynamics of injury, as it can be influenced by other recurrent events. This realization prompted the extension survival models to handle recurrent events.

To address recurrent events, statisticians introduced the Andersen–Gill (AG) model which extends Cox's proportional hazards model by treating each recurrence as a new "starting point" [41]. In the AG model, it is assumed that, under given covariate information, each event is independent and identically distributed, with a shared baseline hazard for all recurrences, so it can explicitly capture the changes in risk probability for subsequent events [42]. Another key assumption of this model is that all event times are also conditionally independent of the given covariates. These assumptions imply that the risks of the recurrent event are identical to those of the first event when all other covariates remain unchanged. When implemented in a sporting context, these assumptions may be unrealistic. For example, Venturelli et al. found that the prior injury significantly elevated the hazard ratio for subsequent re-injury [43]. The Prentice–Williams–Peterson (PWP) model, an extension of Cox's proportional hazards model, can be applied to address this issue. The PWP model considers the order of recurrent events and accounts for the possibility that the baseline hazard of the subsequent event could be higher than that of the previous event [44]. Furthermore, literature identifies several sophisticated extensions for modeling recurrent events, including Wei, Lin, Weissfeld model (WLW) [1], and shared frailty Cox model. The Cox shared frailty model was applied to analyze injury data from an Australian rugby league season [45], and the empirical evidence presented in this study further supports that prior injuries substantially affect the risk of future injury.

Although the above extensions, such as accounting for recurrent events and event ordering, have enhanced analytical approaches in survival analysis, the field has progressed toward the inclusion of unobserved heterogeneity among individuals. It is widely recognized that the physiological status or functional capacity of an individual athlete can exhibit significant variance; however, these factors cannot be directly and easily measured by devices. Therefore, the term "frailty" was introduced into the survival analysis to account for this issue. Frailty can be treated as a random latent and can be used to answer why certain individuals exhibit higher hazard, leading to accelerated failure times than others under identical observed conditions [46]. Incorporating frailty (a random effect) in the model can flexibly capture the inherent heterogeneity of individual profiles [47]. In most models, the unobserved individual-level frailties are assumed to follow gamma or normal distributions [48,49]. Macis applied the frailty model to evaluate the risk factors of injury [50]. He found that the Cox regression model with frailty can provide a better explanation

[1] Wei, Lin, and Weissfeld model (WLW) is another extension of Cox proportional hazards model. In contrast to the AG model, it does not require each recurrence event to have the same baseline hazard.

of the injury data of NBA athletes and emphasized that individual factors can influence the baseline risk of injury [50].

3. The Workload-Injury Model

In this section, we propose a general framework for workload-injury modeling within the context of elite touch rugby athletes, which can be applied to other disciplines in the sports science areas. The framework demonstrates how survival analysis will be utilized to access the relationship between external physical factors and the risk of elite athletes being injured, since external workload metrics (e.g., velocity) directly measure body stress of athletes [25].

As we know, in touch rugby athletes are repeatedly exposed to rapid accelerations, decelerations, and frequent changes of direction, which can cause frequent and sudden increases in external workload, thereby increasing the probability of injury [51]. Based on the characteristics of touch rugby, we consider three states that are non-injury, workload spike, and injury states in our framework. The transitions between these states are illustrated in Figure 1. We assume that the workload spike event repeatedly occurs for each athlete during match periods and once an athlete is injured, he or she does not return to the match. Thus, the workload spike event and injury event can be treated as non-terminal and terminal events, respectively.

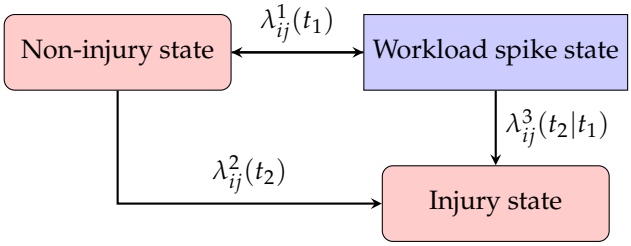


Figure 1. Risk setting in touch rugby match

To further generalize our model, we account for clustering effects and introduce cluster-specific frailty into the model for unobserved heterogeneity between the clusters. Suppose that a group of n athletes can be classified into m clusters based on the common characteristics within the cluster. Let T_{end} denote the end time of the match and T_i^2, for $i = 1, \ldots, n$, denote the time to the injury event. If no injury occurred during the match for the athlete i, then $T_i^2 = \infty$, otherwise $T_i^2 \leq T_{end}$. For the workload spike event, we have n-counting processes $\mathbf{N}^1 = (N_1^1, N_2^1, \ldots, N_n^1)$ adapted to a filtration $\{\mathcal{F}_t^i : t \in [0, T_{end}]\}$, where N_i^1 counts workload spike events for athlete i, during the match period, which are step functions, zero at the start time of the match with jump size 1, such that

$$N_i^1(t) = \sum_{k \geq 1} \mathbf{1}(T_{ik}^1 \leq t),$$

where T_{ik}^1 is the time of the kth spike in the workload event.

Similarly, we also have n-counting processes $\mathbf{N}^2 = (N_1^2, N_2^2, \ldots, N_n^2)$ for the injury event, which is

$$N_i^2(t) = \mathbf{1}(T_i^2 \leq t).$$

Now, we assume that N_i^1 and N_i^2 have random intensities, $\lambda_{ij}^1(t)$ and $\lambda_{ij}^2(t)$, where j denotes the cluster to which the athlete i belongs. Let $\delta_i^1(t)$ and $\delta_i^2(t)$ be indicator functions of the

athlete i who experiences a spike in workload and is injured at time t ($0 \leq t \leq T_{end}$), such that

$$\delta_i^1(t) = \begin{cases} 1, & \text{if workload spike state index} > \text{threshold}, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where a detailed calculation of the workload spike state index is provided in Section 3.1, and

$$\delta_i^2(t) = \begin{cases} 1, & \text{if injured}, \\ 0, & \text{otherwise}. \end{cases} \tag{2}$$

Inspired by the works of Anderson and Gill, Xu et al., and Keiding et al., we propose a general framework of modeling transition rates (hazard rates) illustrated in Figure 1 as,

$$\lambda_{ij}^1(t_1|\mathbf{X}_i(t_1), z_j) = \delta_i^1(t_1)\lambda_{01}\left(t_1 e^{\left\{\boldsymbol{\alpha}_1^\top \mathbf{X}_i(t_1)\right\}}\right) e^{\left\{\beta_{10}+\boldsymbol{\beta}_1^\top \mathbf{X}_i(t_1)+z_j^1\right\}},\ t_1 > 0, \tag{3}$$

$$\lambda_{ij}^2(t_2|\mathbf{X}_i(t_2), z_j) = \delta_i^2(t_2)\lambda_{02}\left(t_2 e^{\left\{\boldsymbol{\alpha}_2^\top \mathbf{X}_i(t_2)\right\}}\right) e^{\left\{\beta_{20}+\boldsymbol{\beta}_2^\top \mathbf{X}_i(t_2)+z_j^2\right\}},\ t_2 > 0, \tag{4}$$

and

$$\lambda_{ij}^3(t_2|t_1, \mathbf{X}_i(t_2), z_j) = \delta_i^3(t_2)\lambda_{03}\left(t_2 e^{\left\{\boldsymbol{\alpha}_3^\top \mathbf{X}_i(t_2)\right\}}\right) e^{\left\{\beta_{30}^\top+\boldsymbol{\beta}_3^\top \mathbf{X}_i(t_2)+z_j^3\right\}},\ 0 < t_1 < t_2, \tag{5}$$

where $\lambda_{ij}^3(t)$ represents the transition rate from a workload spike to injury; $\delta_i^3(t)$ equals 1 if the athlete i is injured after experiencing a spike in workload, otherwise 0; $\mathbf{X}_i(t)$ is a vector of time-dependent covariates, representing factors that contribute to increasing the risk of the athlete i being injured; λ_{01}, λ_{02}, λ_{03} represent the baselines of their corresponding hazard rate functions; and β_{10}, β_{20}, β_{30}, $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$, for $k = 1, 2, 3$, are regression coefficients. Furthermore, z_j^1, z_j^2 and z_j^3 are frailty terms for the cluster j and are assumed to be time-independent and follow a normal distribution ($z_j^d \sim \mathcal{N}(0, \sigma_{dj}^2)$, for $d = 1, 2, 3$). In addition, there is no correlation between frailty terms. It represents an unobserved random effect that captures group-level heterogeneity, as different groups may exhibit distinct movement patterns and therefore affect levels of injury risk. It is also time-independent and shared between all members within the cluster. Moreover, we assume that there is no correlation between clusters. To simplify this, we assume that the frailty terms are independent of covariates and uncorrelated across clusters.

To estimate the parameters of our proposed framework in Equations (3)-(5), we employ the joint calibration method of the Piecewise Exponential Additive Mixed Model (PAMM) framework and the L-BFGS-B optimization algorithm based on the Restricted Maximum Likelihood (REML) objective function and AIC values. In this calibration method, we approximate the hazard function in the PAMM framework. More specifically, this method divides the observation period into small intervals, allowing the model to incorporate covariates that vary over time, such as changes in velocity over time [54]. This is a smart and fast algorithm that works well even when we need to set limits (or boundaries) on the parameters [54,55].

3.1. Workload Spike State Index

As we know, in touch rugby athletes are repeatedly exposed to rapid accelerations, decelerations, and frequent changes of direction, which can cause frequent and sudden increases in external workload, thereby elevating the probability of injury [51]. Based

on these characteristics, the Acute: Chronic Workload Ratio (ACWR) is a good choice to identify the workload spike in touch rugby matches. According to Hulin et al., using metabolic power in the calculation of ACWR can easily detect sudden changes in external workload and provide a robust measurement of risk over short periods based on every person's physical fitness [27]. The metabolic power is an approach that can accurately quantify the instantaneous external workload in matches. In this paper, we adopt the di Prampero et al.'s method to calculate the metabolic power. In this framework, both velocity and acceleration are used to calculate [6]. For detailed calculation processes, see the Appendix B.

Following Hulin et al., the ACWR can be calculated as

$$\text{ACWR} = \frac{\text{Acute Workload}}{\text{Chronic Workload}} \tag{6}$$

where acute workload is short-term load, which usually refers to the most recent workload; chronic workload refers to the average workload over a longer preceding period [27]. In our study, we use ACWR as the workload spike state index in (1). Hulin et al. observe that an ACWR exceeding 1.5 indicates a significant increase in the risk of injury [30]. So, we consider an ACWR that exceeds 1.5 for an athlete as a workload spike event and the threshold in (1) can be set as 1.5 or higher.

3.2. Model for New Zealand Touch Rugby Data

The multiple workload spike events are captured for each athlete in touch rugby matches; however, no injury events were recorded for any athlete throughout the period. This is likely because athletes spend most of their time at relatively low velocities, with only a few instances of reaching very high velocities. They frequently transition between different velocity levels due to the nature of the touch rugby sport. Thus, we consider two key factors that are the velocity variation over the preceding 30 seconds from current time t, denoted as $X_i^A(t)$, and the time of continuous exposure to the workload spike state, denoted as $X_i^B(t)$.

Moreover, in this example, we use the conceptual approach to estimate the ACWR due to the limited information in our data set and, following (6), we have

$$\text{ACWR}(t) = \frac{\overline{\text{MP}}_{30\text{s}}(t)}{\overline{\text{MP}}_{120\text{s}}(t)},$$

where $\overline{\text{MP}}_{30\text{s}}(t)$ represents acute workload at time t, derived from the average metabolic power over the preceding 30 seconds, and $\overline{\text{MP}}_{120\text{s}}(t)$ represents chronic load at time t, derived from the average metabolic power over the preceding 120 seconds. We define a workload spike event when the value of ACWR(t) is greater than 1.5. Thus,

$$\delta_i^1(t) = \begin{cases} 1, & \text{if ACWR(t)} > 1.5, \\ 0, & \text{otherwise,} \end{cases}$$

Based on our proposed framework in this section, the hazard function of this example has the form

$$\begin{aligned} \lambda_{ij}^1(t_1|X_i^A(t_1), X_i^B(t_1), z_j) =& \delta_i^1(t_1)\lambda_{01}\left(t_1 e^{\left\{\alpha_{11}X_i^A(t_1)+\alpha_{12}X_i^B(t_1)\right\}}\right) \\ & \times e^{\left\{\beta_{10}+\beta_{11}X_i^A(t_1)+\beta_{12}X_i^B(t_1)+z_j\right\}}, \end{aligned} \tag{7}$$

where α_{11} and β_{11} are the coefficients of $X_i^A(t)$; and α_{12} and β_{12} are the coefficients of $X_i^B(t)$. Hereafter, we refer to this model (7) as the time scaled frailty model (TSFM).

We also propose the simplified workload-injury models to examine the relationship between workload and injury. First, we build frailty-type models that do not incorporate time-scaled factors. Then, the forms of the models are as follow

$$\lambda_{ij}^1(t_1|X_i^A(t_1), X_i^B(t_1), z_j) = \delta_i^1(t_1)\lambda_{01}(t_1)e^{\{\beta_{10}+\beta_{11}X_i^A(t_1)+\beta_{12}X_i^B(t_1)+z_j\}}, \tag{8}$$

and the model includes only the baseline of the hazard function with cluster-level random effect,

$$\lambda_{ij}^1(t_1|z_j) = \delta_i^1(t_1)\lambda_{01}(t_1)e^{\{\beta_{10}+z_j\}}. \tag{9}$$

In the following, we refer to these two models (8) and (9) as FM and BFM models, respectively.

Second, to evaluate the contribution of cluster-specific frailty, we also proposed further degenerate models that exclude the cluster-specific frailty term and are derived from the well-established Andersen-Gill mode, so they are called AG-type models. The forms of AG-type model are

$$\lambda_{ij}^1(t_1|X_i^A(t_1), X_i^B(t_1)) = \delta_i^1(t_1)\lambda_{01}(t_1)e^{\{\beta_{10}+\beta_{11}X_i^A(t_1)+\beta_{12}X_i^B(t_1)\}}. \tag{10}$$

and the model includes only the baseline of the hazard function,

$$\lambda_{ij}^1(t_1) = \delta_i^1(t_1)\lambda_{01}(t_1)e^{\{\beta_{10}\}}. \tag{11}$$

In the following, we refer to these two models (10) and (11) as AG and BAG models, respectively.

4. Data

In this section, we explore the data from New Zealand touch rugby matches, including male and female matches, apply our general framework to this data set, and investigate how the athletes' movement patterns during matches influence the probability of injury.

The data used for this study consist of high-frequency information collected during touch rugby matches in New Zealand, involving 27 national-level athletes (15 male, 12 female). The dataset captures the velocity, acceleration, and distance at 0.1-second intervals of each athlete during matches, using GPS sensors that operate at a rate of 10-Hz. In this section, we will perform statistical analysis and obtain a deep understanding of each individual's profile.

4.1. Exploratory Data Analysis

Touch rugby is an intermittent sport, characterized by repeated transitions between high- and low-intensity activities, which can be confirmed from Figures 2-5 According to Figures 2 and 4, we can see that female and male athletes run at a velocity between 1 and 0.5 m/s for approximately 70% and 65% of the match time, respectively. At certain moments, both female and male athletes run very fast at more than 4.0 m/s. Furthermore, in Figures 3 and 5, we found that during approximately 80% for females and 70% for males of the match time, the accelerations of the athletes ranged between 0 and 0.1 m/s^2, indicating relatively low acceleration levels. So we can conclude that athletes are in low-intensity activity states during the majority of match time.

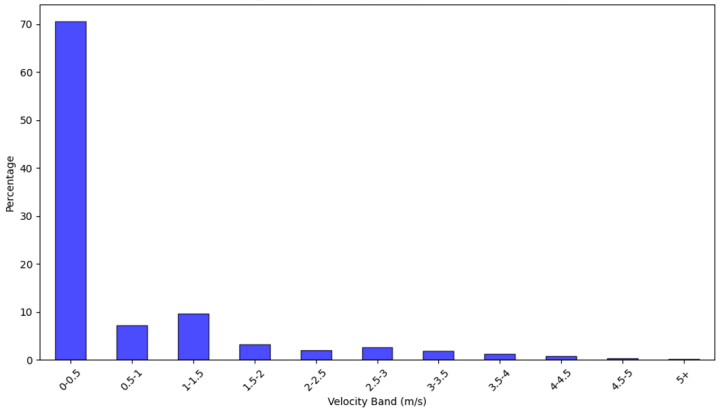


Figure 2. Velocity distribution for 12 female athletes during the matches

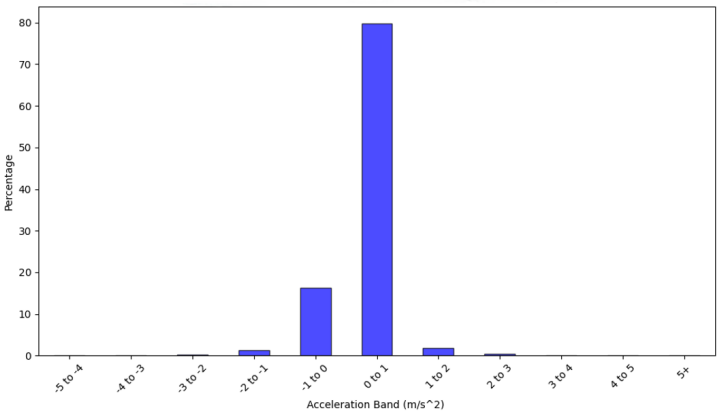


Figure 3. Acceleration distribution 12 female athletes during the matches

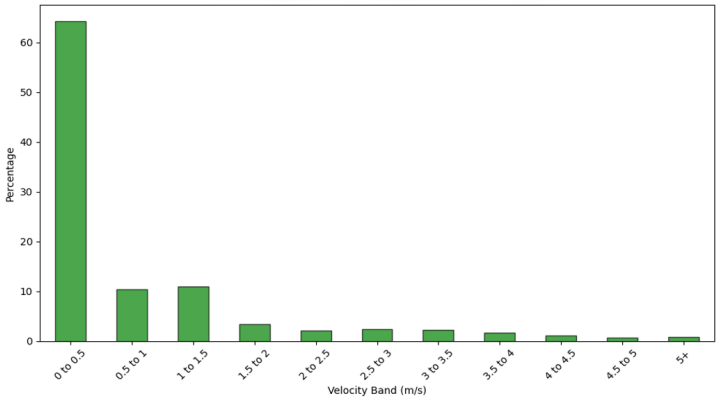


Figure 4. Velocity distribution for 15 male athletes during the matches

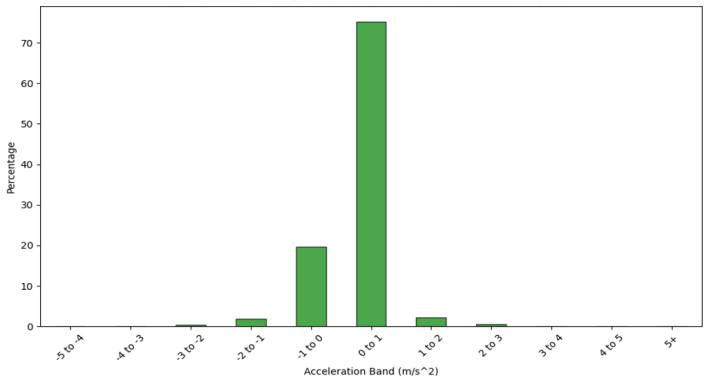


Figure 5. Acceleration distribution for 15 male athletes during the matches

Studies show that velocity and acceleration are the key components of external workload [6,56]. In particular, high-intensity movements are the main factor that contribute to athlete's fatigue and increase the risk of injury [56]. Thus, given the nature of touch rugby, it is necessary to segment in-play and rest periods during matches, which helps to understand workload, recovery, and overall performance patterns, and formed the basis for further modeling. Dwyer and Gabbett show that walking speeds in the sport field are around 1.0 m/s for female and 1.2 m/s for male [57]. So, we define rest periods for male athletes as the interval in which their velocities are below 1.2 m/s and a threshold of 1.0 m/s was used for female. In other to more accurately identify the patterns of rest and in-play periods, we use moving average technique to smooth short-term (3 seconds) fluctuations, which is followed by Marutani et al.'s research. After that, we apply the predefined thresholds to perform the segmentation. In addition, following Marutani et al., we introduce an additional condition to the definition of the in-play segment [58]. Specifically, this refers to the period in which the athlete's velocity exceeds the predefined threshold for more than 5 seconds [58].

The segmentations of the period for one female and one male athlete are shown in Figures 6 and 8. The segmentations for the remaining athletes are presented in the Appendix A. The gray and white areas in these figures represent the rest and in-play periods, respectively. The blue lines represent the 3 seconds moving average velocity traces. According to these figures, each athlete exhibits multiple velocity spikes intermittently throughout the entire matches. These results form the foundation for the subsequent analysis and modeling of the probability of developing fatigue.

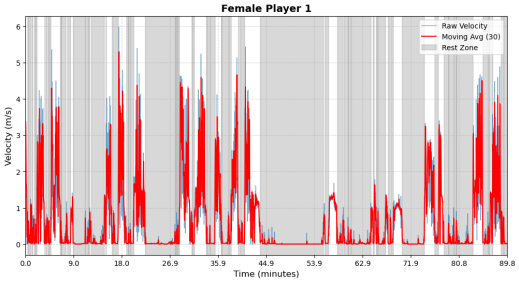


Figure 6. Female athlete 1 - Period Segmentation

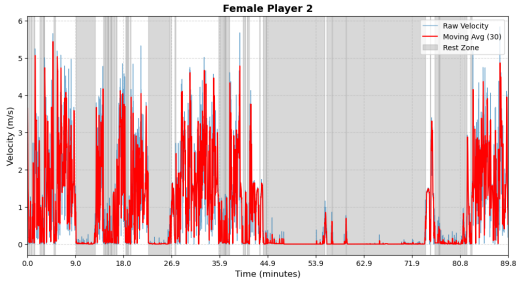


Figure 7. Female athlete 2 - Period Segmentation

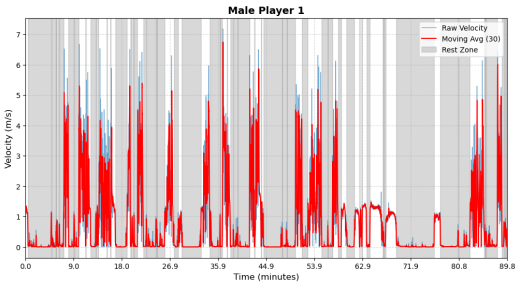


Figure 8. Male athlete 1 - Period Segmentation

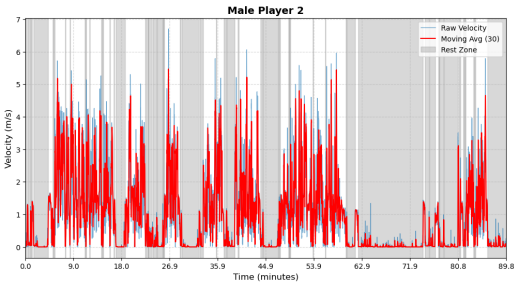


Figure 9. Male athlete 2 - Period Segmentation

4.2. Athletes' Profiles Analysis

To gain a deeper understanding of the differences among athletes' movement patterns, we apply the Gaussian Mixture Model (GMMs) to cluster velocity data taken during in-play periods into high-intensity and low-intensity phases for each athlete. The definition of in-play state refers to Subsection 4.1. Furthermore, in order to examine whether there is a cluster effect within the teams, we use K-means to cluster female and male athletes into 3 and 4 groups, respectively, based on the estimated results from GMMs. Detailed information is shown in Tables 1 and 2 below.

Table 1. Summary statistics for female athletes by clusters

Cluster	Athlete	μ_{low}	σ_{low}	w_{low}	μ_{high}	σ_{high}	w_{high}	In-Play (%)	Rest (%)	Total Distance
0	FP 2	1.25	0.51	0.66	3.05	0.80	0.34	27.25	72.75	3253.70
0	FP 3	1.27	0.49	0.58	3.09	0.79	0.42	33.81	66.19	4116.50
0	FP 10	1.12	0.61	0.71	3.13	0.87	0.29	26.66	73.34	2914.30
0	FP 4	1.21	0.50	0.51	3.23	0.81	0.49	16.46	83.54	2319.60
0	FP 11	1.29	0.54	0.67	3.26	0.94	0.33	22.53	77.47	2808.30
0	FP 5	1.14	0.47	0.66	3.39	0.95	0.34	22.81	77.19	2734.40
1	FP 9	1.20	0.18	0.48	2.58	1.01	0.52	23.42	76.58	2955.00
1	FP 1	1.22	0.23	0.46	2.60	1.18	0.54	22.91	77.09	2978.10
1	FP 12	1.29	0.24	0.46	2.74	1.24	0.54	20.96	79.04	2724.50
1	FP 6	1.19	0.22	0.50	2.80	1.23	0.50	23.41	76.59	2912.40
2	FP 7	1.21	0.35	0.60	3.30	0.69	0.40	32.46	67.54	3881.50
2	FP 8	1.14	0.47	0.48	2.75	0.77	0.52	15.70	84.30	2224.90

First, parameters are estimated from a Gaussian Mixture Model (GMM) with two-component fitted to each female athlete's velocity distribution (m/s). The two components represent the low- and high-intensity phases. μ_{low}, σ_{low}, w_{low} refer to average velocity (m/s), standard deviation (m/s), and proportion in low-intensity phase, respectively. μ_{high}, σ_{high}, w_{high} refer to average velocity (m/s), standard deviation (m/s), and proportion in high-intensity phase, respectively. These calculations are only based on the data from in-play periods. Second, in-play (%) and rest (%) represent the fraction of total match time that each athlete spends in the in-play and recovery phase, respectively. Last, total distance refers to the total distance covered of each athlete during the whole data collection period.

From Table 1, we can see that each female cluster exhibits different movement patterns. First, Cluster 0 is characterized by a relatively high average velocity (μ_{high}) during the high-intensity phase, and the proportion of low-intensity activity is higher than that of high-intensity activity. Second, Cluster 1 is characterized by low values of σ_{low}, suggesting more consistent movement patterns with less variability in the low speed components. This group also maintains balanced weights (w_{low}, w_{high}), indicating even contributions across intensity levels. Additionally, compared to other clusters, they exhibit a higher σ_{high}, indicating a greater variability in their speed during the high-intensity phase. Third, the athletes demonstrate moderate average speed for both low- and high-intensity states as well as moderate variability in their speed patterns in Cluster 2.

Table 2. Summary statistics for male athletes by clusters

Claster	Athlete	μ_{low}	σ_{low}	w_{low}	μ_{high}	σ_{high}	w_{high}	In-Play (%)	Rest (%)	Total Distance
0	MP 14	1.43	0.64	0.61	3.46	1.027	0.393	27.90	72.10	4204.30
0	MP 5	1.32	0.76	0.57	3.78	1.050	0.427	14.46	85.54	3066.30
0	MP 7	1.55	0.70	0.61	3.47	1.144	0.390	21.30	78.70	3367.10
0	MP 6	1.66	0.73	0.52	3.69	0.987	0.476	17.42	82.58	3348.50
1	MP 12	1.11	0.75	0.56	3.43	1.210	0.444	16.68	83.32	2655.70
1	MP 1	1.27	0.48	0.58	3.40	1.065	0.424	25.96	74.04	3849.70
1	MP 9	1.27	0.54	0.56	3.35	1.212	0.438	17.28	82.72	3085.30
1	MP 4	1.24	0.40	0.47	3.46	1.036	0.528	21.43	78.57	3839.80
1	MP 15	1.21	0.42	0.41	3.28	1.250	0.595	15.88	84.12	2841.00
2	MP 2	1.20	0.67	0.66	3.26	0.810	0.343	37.33	62.67	4818.20
2	MP 13	1.34	0.55	0.59	3.29	0.856	0.411	27.31	72.69	4203.50
3	MP 10	1.31	0.36	0.50	2.98	1.173	0.498	29.63	70.37	4450.70
3	MP 3	1.34	0.16	0.37	2.92	1.273	0.628	23.90	76.10	3864.50
3	MP 8	1.31	0.14	0.40	3.06	1.268	0.603	20.11	79.89	3722.60
3	MP 11	1.40	0.23	0.33	3.02	1.232	0.669	17.62	82.38	3211.50

First, parameters are estimated from a Gaussian Mixture Model (GMM) with two-component fitted to each male athlete's velocity distribution (m/s). The two components represent the low- and high-intensity phases. μ_{low}, σ_{low}, w_{low} refer to average velocity (m/s), standard deviation (m/s), and proportion in low-intensity phase, respectively. μ_{high}, σ_{high}, w_{high} refer to average velocity (m/s), standard deviation (m/s), and proportion in high-intensity phase, respectively. These calculations are only based on the data from in-play periods. Second, in-play (%) and rest (%) represent the fraction of total match time that each athlete spends in the in-play and recovery phase, respectively. Last, total distance refers to the total distance covered of each athletes during the whole data collection period.

According to Table 2, male athletes also exhibit different movement patterns for the different clusters. First, Cluster 0 is characterized by the highest average speed during the in-play period, along with the highest speed standard deviations in the low-intensity state, indicating a tendency toward higher and more volatile speeds during the in-play periods. Second, Cluster 1 shows moderate mean velocities with mid-range variation suggesting more balanced movement patterns. Third, Cluster 2 athletes exhibit moderate average speed, combined with lowest σ_{high}, indicating more stable high-intensity running efforts. In contrast, Cluster 3 is distinguished by the lowest speed in the high-intensity state with the highest standard deviation and weight. This implies that these athletes consistently maintain high-speed states and the highest fluctuation.

5. Results and Discussion

This section presents the model estimation results and model performance comparison, followed by a post-model analysis. The key findings are then examined and discussed in the context of the research objectives to highlight their practical implications.

5.1. Model Estimation and Evaluation

In this section, we interpret the parameters following the estimation procedure in Section 3 to investigate the influence of the velocity variation and the time of continuous exposure to the workload spike state on the risk of injury.

Table 3 shows the results of the estimation of the parameters in Equations (7)-(11) based on the data described in Section 4. The relatively small standard deviations for most parameters indicate that the parameter has been accurately estimated, with the exception of σ_z. From Table 3, we find that both the velocity variation in the preceding 30 seconds and the time of continuous exposure to the workload spike state are significant predictors of the risk of injury. As we know, athletes experience a workload spike state, the risk of injury increases significantly [30]. The estimated results for all proposed models show

Table 3. Parameter Estimates

	Model	α_{11}	α_{12}	β_{10}	β_{11}	β_{12}	σ_z
Male	BAG	–	–	1.038*	–	–	–
		–	–	(0.002)	–	–	–
	BFM	–	–	1.004*	–	–	0.133*
		–	–	(0.067)	–	–	(0.233)
	AG-SDV	–	–	0.470*	0.923*	–	–
		–	–	(0.003)	(0.004)	–	–
	FM-SDV	–	–	0.447*	0.916*	–	0.104*
		–	–	(0.052)	(0.004)	–	(0.181)
	TSFM-SDV	-29.512*	–	0.542*	0.651*	–	0.124*
		(0.001)	–	(0.0624)	(0.004)	–	(0.220)
	AG-D	–	–	-0.231*	–	1.494*	–
		–	–	(0.004)	–	(0.003)	–
	FM-D	–	–	-0.245*	–	1.491*	0.065*
		–	–	(0.033)	–	(0.003)	(0.116)
	TSFM-D	–	-32.478*	0.310*	–	0.099*	0.111*
		–	(3.341)	(0.056)	–	(0.005)	(0.196)
	AG	–	–	-0.393*	0.442*	1.363*	–
		–	–	(0.005)	(0.004)	(0.003)	–
	FM	–	–	-0.401*	0.439*	1.361*	0.045*
		–	–	(0.023)	(0.004)	(0.003)	(0.020)
	TSFM	-0.343*	-16.576*	0.267*	0.194*	0.025*	0.086*
		(0.001)	(0.301)	(0.044)	(0.004)	(0.005)	(0.153)
Female	BAG	–	–	1.057*	–	–	–
		–	–	(0.002)	–	–	–
	BFM	–	–	1.040*	–	–	0.058*
		–	–	(0.033)	–	–	(0.131)
	AG-SDV	–	–	0.440*	1.273*	–	–
		–	–	(0.004)	(0.005)	–	–
	FM-SDV	–	–	0.439*	1.273*	–	0.011*
		–	–	(0.007)	(0.005)	–	(0.031)
	TSFM-SDV	-0.047*	–	0.422*	1.312*	–	0.007*
		(0.000)	–	(0.005)	(0.006)	–	(0.026)
	AG-D	–	–	-0.236*	–	2.245*	–
		–	–	(0.005)	–	(0.005)	–
	FM-D	–	–	-0.236*	–	2.245*	0.017*
		–	–	(0.011)	–	(0.005)	(0.041)
	TSFM-D	–	-35.001*	0.343*	–	0.215*	0.010*
		–	(0.003)	(0.008)	–	(0.008)	(0.039)
	AG	–	–	-0.366*	0.538*	2.013*	–
		–	–	(0.005)	(0.006)	(0.006)	–
	FM	–	–	-0.362*	0.540*	2.014*	0.020*
		–	–	(0.013)	(0.006)	(0.006)	(0.048)
	TSFM	-0.394*	-28.055*	0.318*	0.120*	0.165*	0.042*
		(0.000)	(0.004)	(0.025)	(0.005)	(0.008)	(0.099)

Values in parentheses indicate standard deviation. * indicates that all parameter estimates are statistically significant at the $p < 0.001$ level. BAG refer to the model that allows for recurrent events but doesn't include the subject specific frailty and covariates. The model name with SDV refers to the model which only considers the velocity variation over the preceding 30s. The model name with D refers to the model which only considers the continuous duration spent in the high-intensity state.

that the greater changes in velocity in the preceding 30 seconds increase the probability of injury. For example, in the FM model, if the other conditions remain unchanged, then a unit increase in the standard deviation of the velocity is associated with a 55.12% which equals $e^{0.439} - 1$ increase for male athletes and 71.60% which equals $e^{0.540} - 1$ increase for female athletes, respectively, in the hazard rate of entering the workload spike state.

Moreover, based on the estimated results for our most complex model (TSFM model), α_{11} and α_{12} in the TSFM model for both male and female are estimated to be negative, indicating that an increase in the velocity variation in the preceding 30 seconds and continuous exposure to the workload spike state will speed up to the point of injury. More specifically, the factor of continuous exposure to the workload spike state increases injury risk more than the factor of velocity variation over 30 seconds and leads the individual athlete toward the workload spike state sooner in the touch rugby match. By comparing the estimated values of α_{11} and α_{12} between the TSFM model for male and female athletes, we can also observe that female athletes tend to experience fatigue and injury faster than male athletes under the condition of increasing the variation of velocity and sustaining high-intensity activities.

Table 4. Model comparison for male and female athletes

Group	Model	AIC	BIC	Deviance Explained	REML
Male	BAG	1,006,369	1,006,790	0.020	-503,300
	BFM	1,004,847	1,005,303	0.228	-502,550
	AG-SDV	944,279	944,712	0.130	-472,260
	FM-SDV	943,411	943,878	0.132	-471,840
	TSFM-SDV	939,015	939,482	0.140	-469,660
	AG-D	798,830	799,260	0.389	-399,510
	FM-D	798,469	798,933	0.389	-399,340
	TSFM-D	705,680	706,147	0.554	-352,980
	AG	788,351	788,793	0.407	-394,280
	FM	788,187	788,664	0.408	-394,200
	TSFM	702,079	702,558	0.561	-351,180
Female	BAG	802,058	802371	0.023	-401,110
	BFM	801,833	802169	0.031	-401,000
	AG-SDV	741,056	741,381	0.158	-370,630
	FM-SDV	741,046	741,393	0.159	-370,620
	TSFM-SDV	740,254	740,600	0.160	-370,230
	AG-D	633,661	633,985	0.398	-316,910
	FM-D	633,627	633,974	0.398	-316,900
	TSFM-D	560,273	560,618	0.561	-280,230
	AG	626,511	626,847	0.414	-313,340
	FM	626,458	626,816	0.414	-313,320
	TSFM	556,615	556,974	0.569	-278,420

REML refers to Restricted Maximum Likelihood. This method usually is used for the mixed-effects and additive models with less biased for variance [59]. AIC and BIC refer to Akaike Information Criterion and Bayesian Information Criterion, respectively.

Finally, table 4 shows the evaluation of the performance of the proposed model using AIC, BIC, deviance explained, and REML. Three key observations can be drawn from this table. First, Models incorporating covariates perform better than those without covariates. This indicates that the likelihood of injury is strongly related to velocity variations and the time of continuous exposure to the workload spike state for both female and male athletes. Second, The hazard rates exhibit a cluster-level effect, as models incorporating a cluster-specific frailty term show lower AIC, and BIC values, as well as a higher deviance

explained rate and REML compared to models without the frailty term, while all other settings remain the same. This further demonstrates that members within the same cluster share unobserved risk factors and that differences in movement patterns lead to varying levels of risk of injury. Third, the TSFM model consistently outperforms all other proposed models, which imply both velocity fluctuations and sustained duration of workload spike state impact on the time to injury.

5.2. Post-Estimation Analysis

This section examines the baseline hazard, the marginal effects of the most influential factor, and the analysis of heterogeneity between clusters. Furthermore, we elucidate cluster-specific frailty effects in touch rugby matches.

5.2.1. Baseline Hazard Analysis

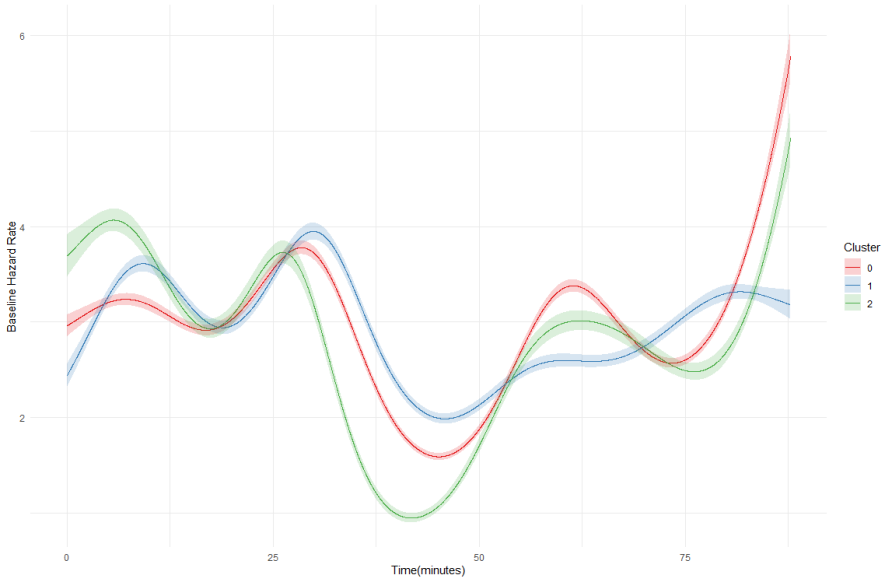


Figure 10. Cluster-specific baseline hazard estimates for female athletes

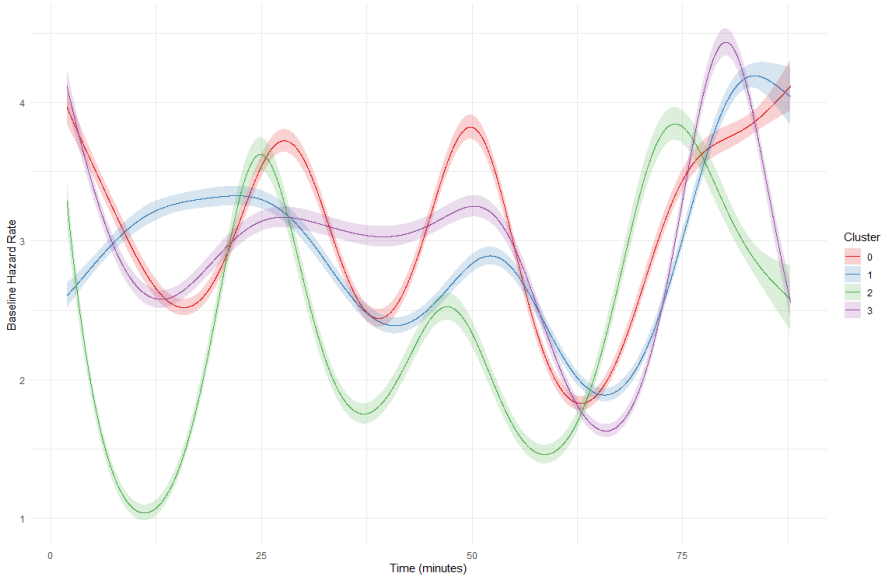
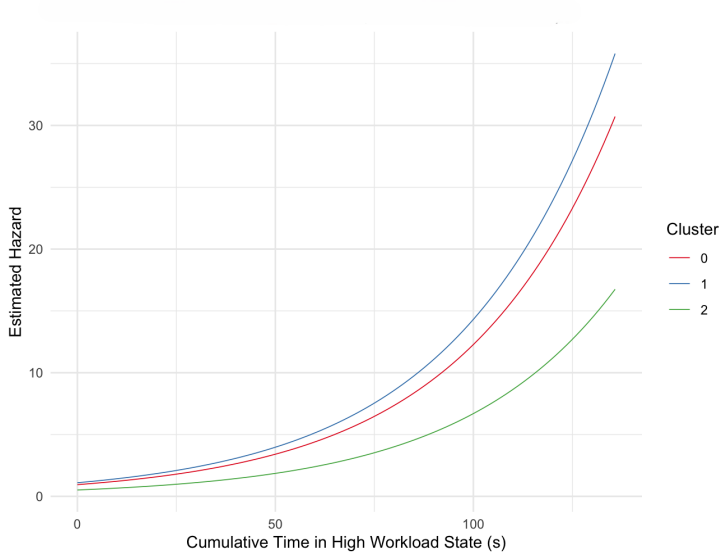


Figure 11. Cluster-specific baseline hazard estimates for male athletes

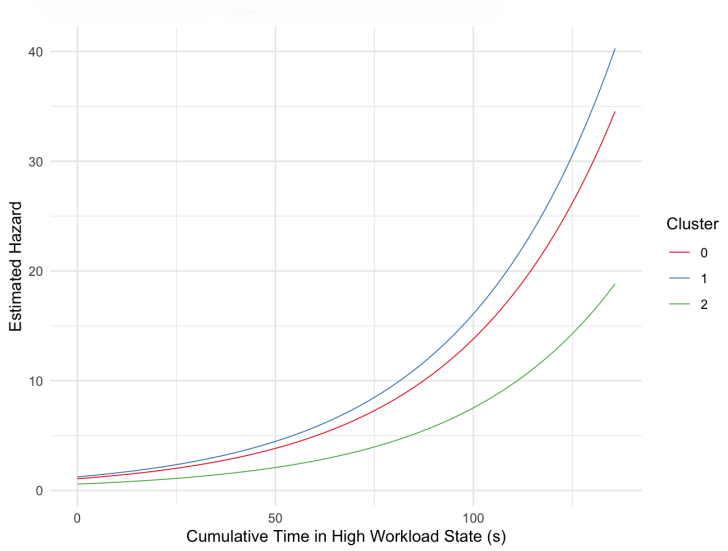
The baseline hazard rate reflects the instantaneous injury risk regardless of the individual athlete's movement behavior, which is the foundation of our proposed models. Figures 10 and 11 illustrate the estimated baseline hazard line stratified by cluster for female and male athletes, respectively. The results highlight that baseline hazards fluctuate during matches, exhibiting distinct temporal hazard patterns between clusters. Regarding female athletes, from Figure 10, we observe a significant increase in the underlying hazard at the end of the match for all clusters, suggesting that cumulative fatigue significantly increases the risk of injury. Three spikes are identified on the baseline hazard line for Clusters 0 and 2 during the interval 7-8 minutes, 26-28 minutes, and 50-52 minutes, indicating a higher probability of injury occurrence. In addition, Cluster 1 maintained relatively stable baseline hazard rates in contrast to those of the other two clusters. Regarding male athletes, the hazard trajectories were more varied. All clusters showed a late-stage surge in baseline hazard rates, peaking after 70 minutes, indicating fatigue-related risks. Cluster 1 exhibits an increase in baseline hazard rates in the early stage of the match, which might imply inadequate warm up exercises. These baseline hazard trends support the hypothesis that workload patterns influence injury risk differently across clusters throughout the entire observation window.

5.2.2. Partial Effect Analysis

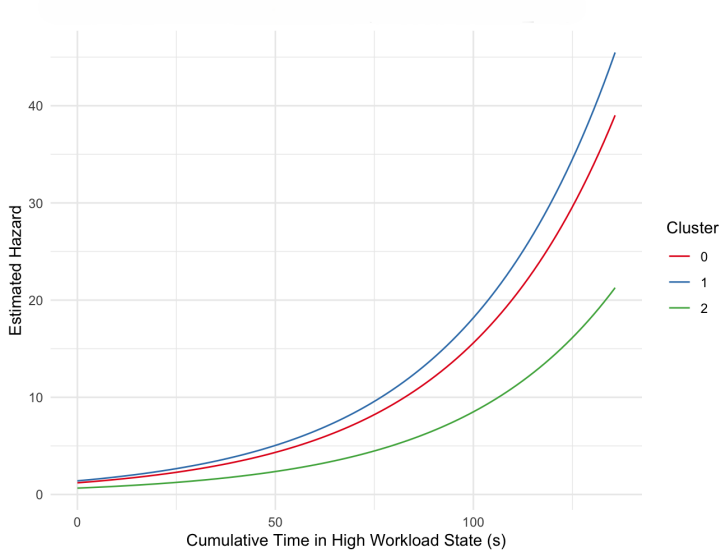
Based on the analysis of the baseline hazard rates, we now examine the partial effect of the time of continuous exposure to the workload spike state to understand how this factor impacts the underlying risks. We specifically focus on the partial effect of this factor, as it plays a more crucial role in our workload-injury model (TSFM) according to the results in Table 4. Figures 12 and 13 illustrate the estimated partial effects of continuous time exposure to the workload spike state on the hazard function (TSFM model) by cluster at fixed levels of velocity variation (25th, 50th and 75th percentiles). The hazard rates in Figures 12 and 13 exhibit non-linear, accelerating relationships with cumulative high-workload duration across all clusters. Furthermore, they show that the hazard rate increases more rapidly with greater variation in velocity for both male and female athletes.



(a) 25th percentile

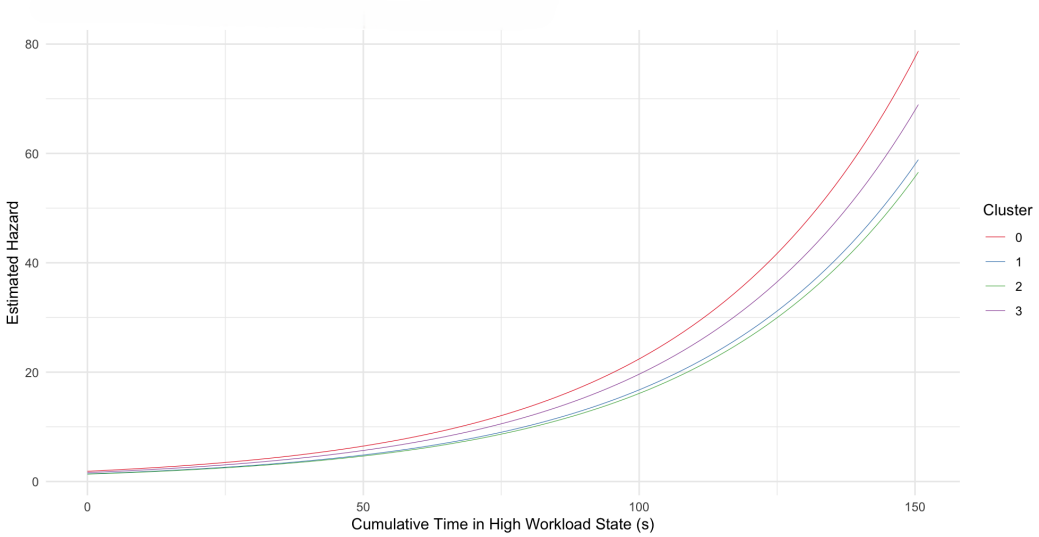


(b) 50th percentile

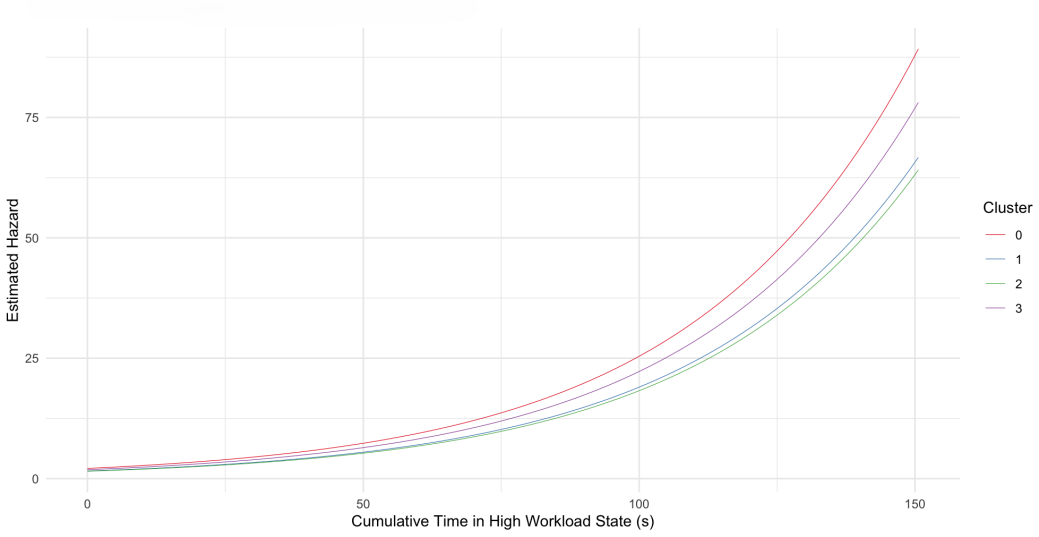


(c) 75th percentile

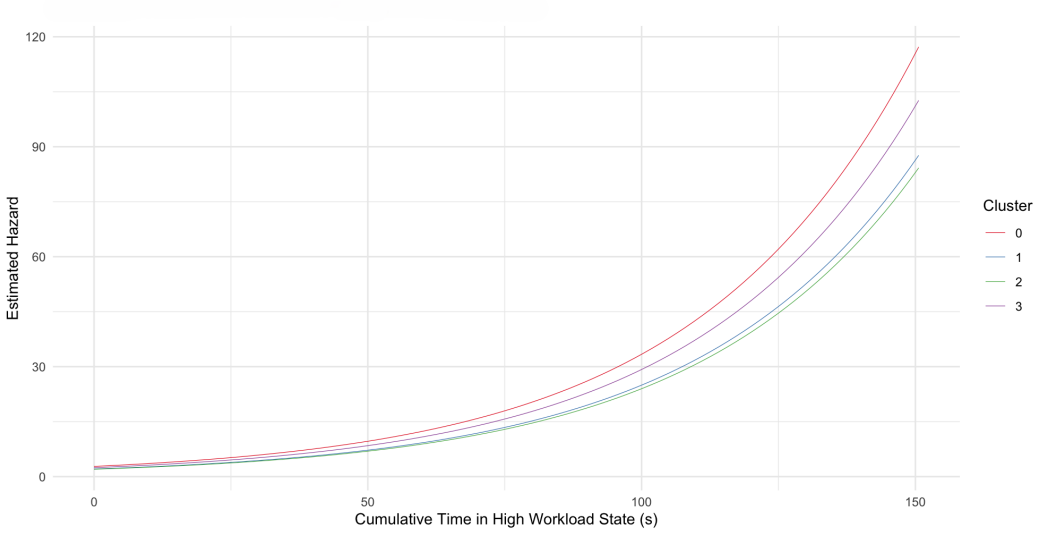
Figure 12. Partial effects of cumulative high-workload duration on hazard for female athletes across different velocity variation levels



(a) 25th percentile



(b) 50th percentile



(c) 75th percentile

Figure 13. Partial effects of cumulative high-workload duration on hazard for male athletes across different velocity variation levels

Having discussed the common features across all the clusters, we now compare the differences among the clusters. For female athletes, Cluster 2 is associated with the lowest estimated hazard, while Cluster 1 consistently demonstrates the highest risk at different levels of velocity variation. Combined with the observations in Table 1, we find that the hazard rate is relatively lower when movement patterns exhibit smaller differences in the standard deviation of velocity between the low- and high-intensity phases, as well as a lower overall velocity variation. For male athletes, we find a trend similar to that observed among female athletes.

5.2.3. Heterogeneity Analysis between Clusters

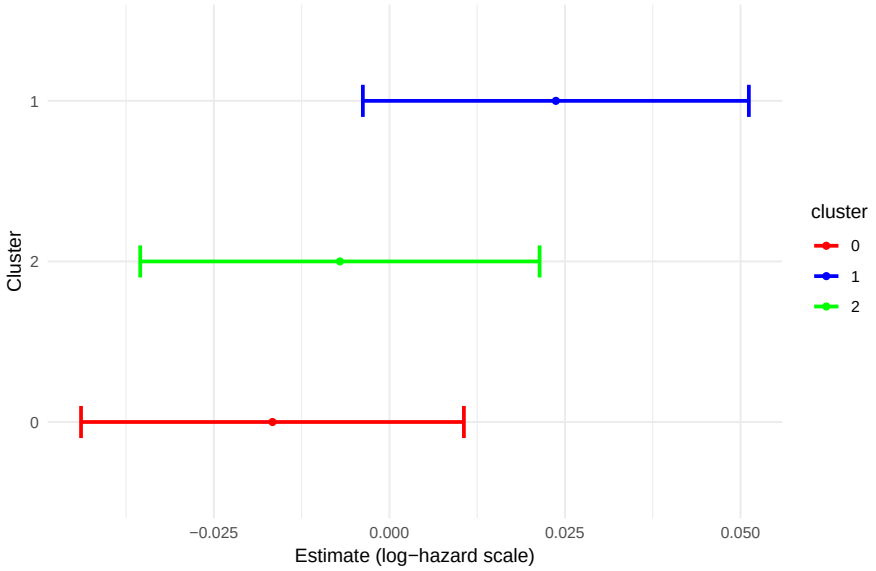


Figure 14. Random effects (frailty) estimates for female clusters with 95% confidence intervals

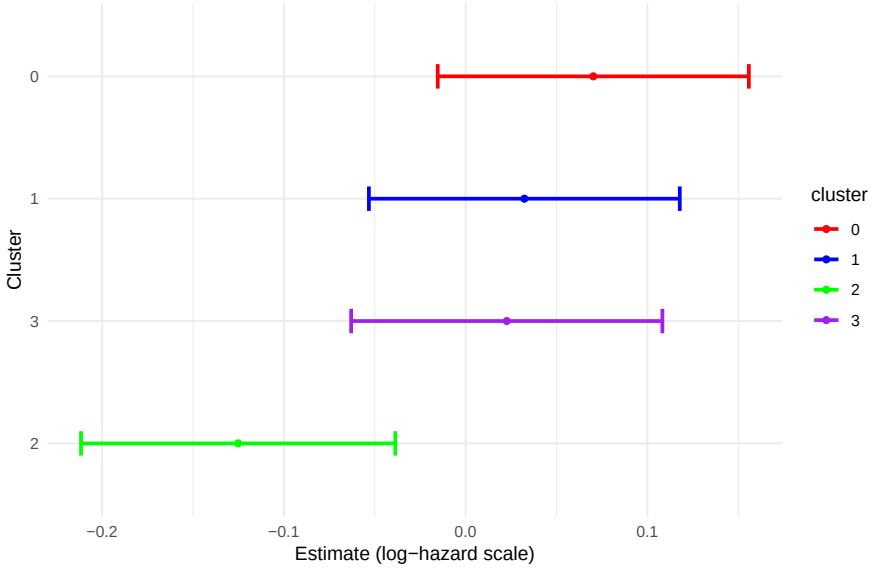


Figure 15. Random effects (frailty) estimates for male clusters with 95% confidence intervals

Figures 14 and 15 present the cluster-specific frailty term (random effects) with 95% confidence intervals, which is used to quantify an unobserved heterogeneity between

clusters. Among female athletes, Cluster 1 demonstrated a positive frailty estimate that exceeded all other clusters, indicating a higher baseline risk compared to all other clusters. According to the information in Figure 1, Cluster 1 is characterized by velocity variability. Specifically, the standard deviations of the velocities exceeds 1 m/s during high-intensity phases (approximately $> 39\%$ of the average velocities), which are higher than all other clusters. In addition, high-intensity activity constitutes at least half of the active in-play duration. Furthermore, the frailty terms for Clusters 0 and 2 are estimated to be negative, suggesting that they have lower risks than the baseline risk. Clusters 0 and 2 demonstrate reduced velocity fluctuations during the high-intensity phase relative to Cluster 1.

In male athletes, positive frailty estimates were observed for Clusters 0, 1, and 3, with Cluster 0 exhibiting the highest value, identifying Cluster 0 as the highest-risk group. In contrast, the estimated frailty term of Cluster 2 is negative, thereby we can conclude that this cluster has the lowest likelihood of getting fragile and injured. The distinguishing movement characteristic of Cluster 2, which differentiates it from the other cohorts, is that the standard deviations of the velocity during the high-intensity phases are lower than those of the others, with a value consistently below 0.9 m/s ($< 26.1\%$ of the average velocities).

In summary, baseline risks are varied between clusters due to different movement patterns. Variation in velocity during high-intensity exercise has a significant impact on the baseline risk. Greater movement fluctuations in velocity appear to significantly predispose athletes to injury in the touch rugby matches, as reflected by the higher frailty estimates.

6. Conclusion

Survival models have been widely applied in the field of sports analytics to predict injury. In this study, we extend the traditional one-dimensional survival model to a more general workload-injury framework, which incorporates the transition between 3 states: non-injury, workload spike, and injury states. We allow for mutual transit between the non-injury and workload spike states. However, we do not allow for the transition from the injury state to the other two states. These designs are motivated by our empirical observations and the characteristics of the movement pattern in touch rugby matches. The key pattern observed is that athletes frequently transit between high- and low-intensity activities during matches.

The proposed Time-Scaled Frailty Model (TSFM) incorporates three key components to better assess injury and workload spike risk in touch rugby: a time-scaling mechanism to reflect how covariates (e.g., velocity variation) accelerate or decelerate time to overload or injury; a normally distributed frailty term to capture cluster effects and unobserved heterogeneity (with potential extensions to correlated or alternative frailty distributions such as gamma); and time-varying covariates to model dynamic changes in risk factors. Applied to kinematic time-series data from touch rugby matches, the model shows that short-term velocity fluctuations and cumulative high-intensity activity significantly influence time to workload spikes and injury states. Its superior performance over competing models underscores the importance of accounting for both observed covariates and unobserved cluster-level differences, revealing that overload risk and baseline hazard patterns vary across clusters. These findings highlight the need for personalized workload monitoring and demonstrate that survival models integrating time-scaling and frailty components more accurately capture fatigue accumulation and injury risk dynamics in touch rugby.

From a practical perspective, there are three key pieces of information. First, keeping track of short term variability can help identify early signal for overloading. Second, setting a limit on how long athletes can remain continuously in the high workload state is crucial for managing the total workload athletes have to take. Third, adjusting training

and substitution strategies based on clusters can help improve recovery planning and reduce injury risk. This study offers a robust framework for identifying risk factors for injury and their underlying dynamics. It enables the development of evidence-based and individual-specific training and competition strategies tailored to the specific demands of different sporting activities.

However, this study also has certain limitations, as there is no universally accepted standard definition of an overloading or workload spike event. In this study, an ACWR greater than 1.5 is used as the threshold to define a workload spike. It may also be defined using a different approach based on your research purpose. Tactical roles and match contexts were not included in this study, and data came from New Zealand touch rugby matches. Future work should consider linking these models with injury records and include tactical factors.

Author Contributions: Conceptualization, S.S. and T.H.; methodology, S.S.; software, T.H. and S.S.; resources, K.S.; data curation, K.S. and T.H.; writing—original draft preparation, T.H. and S.S.; writing—review and editing, N.W., S.S. and K.S.; visualization, T.H. and S.S.; supervision, S.S. and N.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: The data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: GenAI has been used for grammar checking.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACWR	Acute:Chronic Workload Ratio
AG	Andersen-Gill Model
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
BAG	Baseline Andersen-Gill Model
BIC	Bayesian Information Criterion
BFM	Baseline Frailty Model
COM	Center of Mass
Cox PH	Cox Proportional Hazards Model
ES	Equivalent Slope
FM	Frailty Model
GMM	Gaussian Mixture Model
GPS	Global Positioning System
Hz	Hertz
K-means	K-means clustering
L-BFGS-B	Limited-memory Broyden–Fletcher–Goldfarb–Shanno with Boundaries
MP	Metabolic Power
PAMM	Piecewise Exponential Additive Mixed Model
PWP	Prentice–Williams–Peterson Model
REML	Restricted Maximum Likelihood
RPE	Rating of Perceived Exertion
SDV	Velocity variation (standard deviation of velocity) Model
sRPE	Session Rating of Perceived Exertion
TSFM	Time-Scaled Frailty Model
WLW	Wei-Lin-Weissfeld Model

Appendix A. Segmentation of Rest vs. In-play Periods

The segmentation of the period for the remaining athletes in Section 4.1 is listed below.

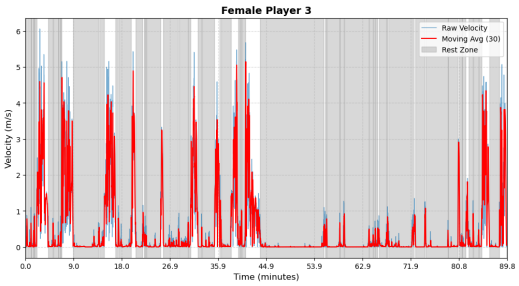


Figure A1. Female athlete 3 - Period Segmentation

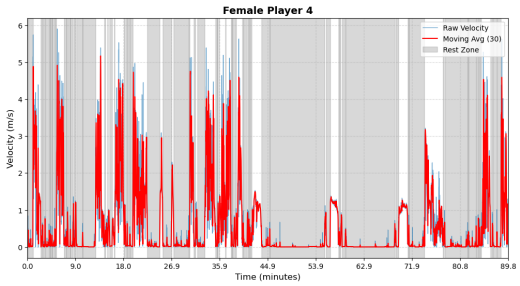


Figure A2. Female athlete 4 - Period Segmentation

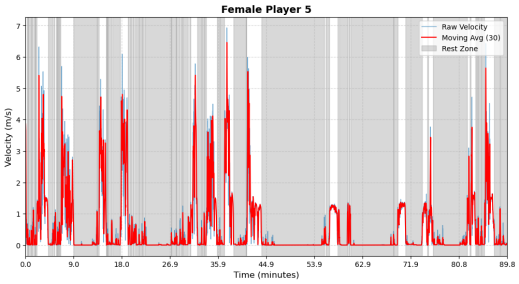


Figure A3. Female athlete 5 - Period Segmentation

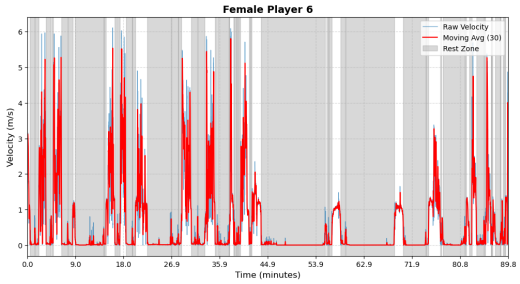


Figure A4. Female athlete 6 - Period Segmentation

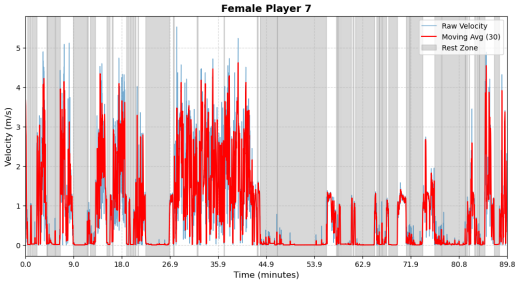


Figure A5. Female athlete 7 - Period Segmentation

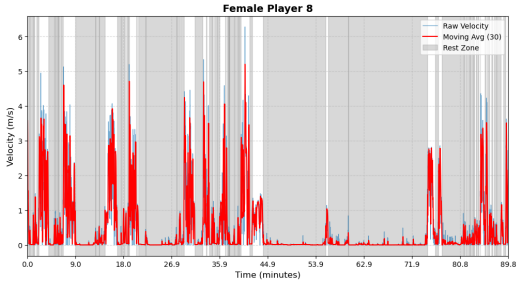


Figure A6. Female athlete 8 - Period Segmentation

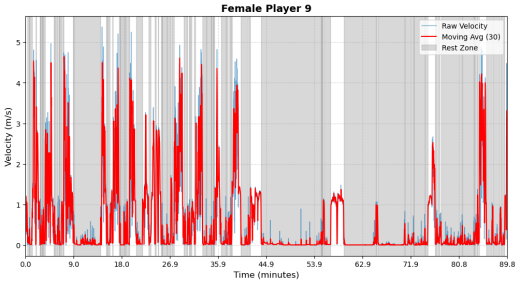


Figure A7. Female athlete 9 - Period Segmentation

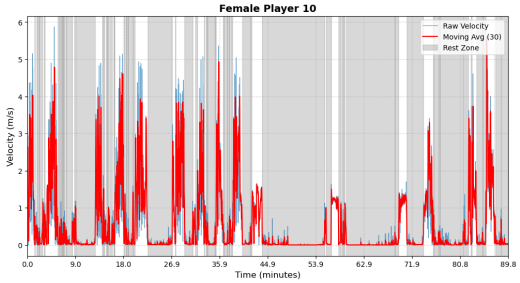


Figure A8. Female athlete 10 - Period Segmentation

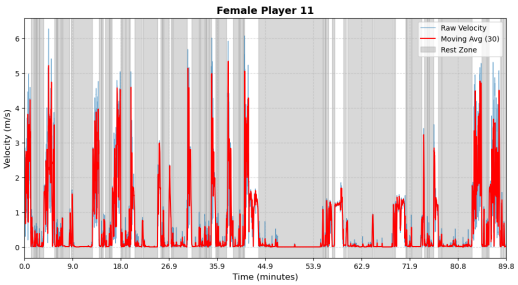


Figure A9. Female athlete 11 - Period Segmentation

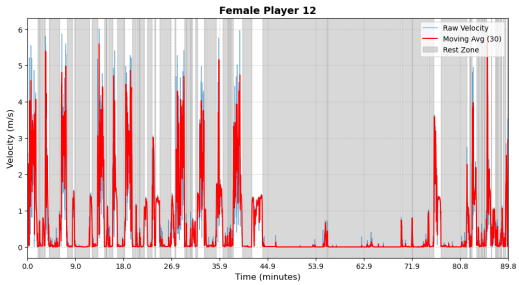


Figure A10. Female athlete 12 - Period Segmentation

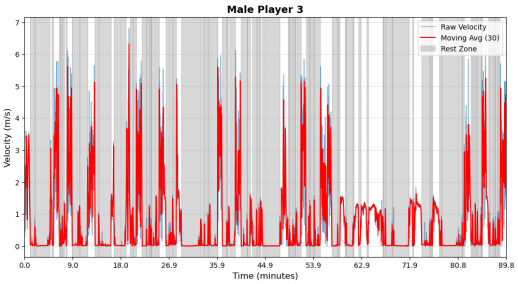


Figure A11. Male athlete 3 - Period Segmentation

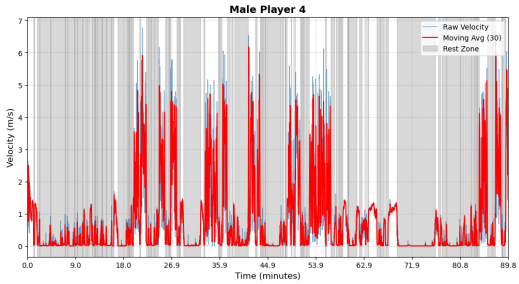


Figure A12. Male athlete 4 - Period Segmentation

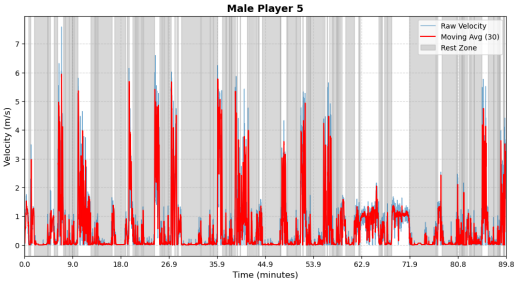


Figure A13. Male athlete 5 - Period Segmentation

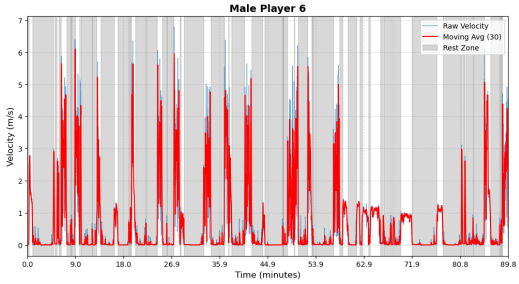


Figure A14. Male athlete 6 - Period Segmentation

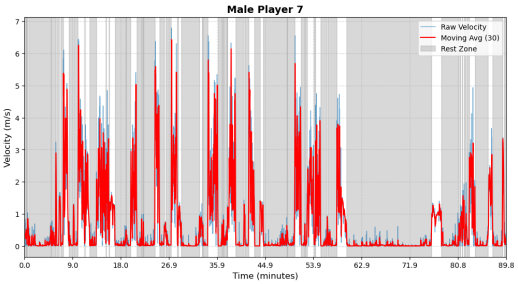


Figure A15. Male athlete 7 - Period Segmentation

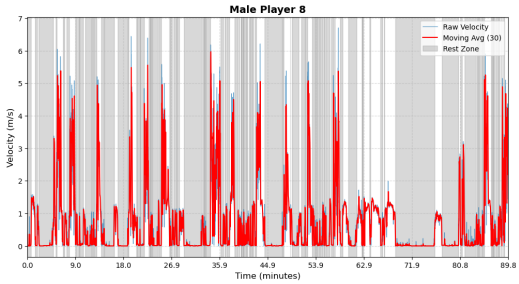


Figure A16. Male athlete 8 - Period Segmentation

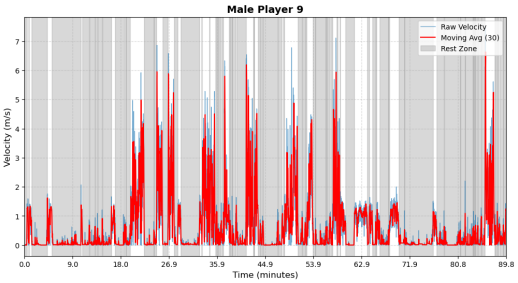


Figure A17. Male athlete 9 - Period Segmentation

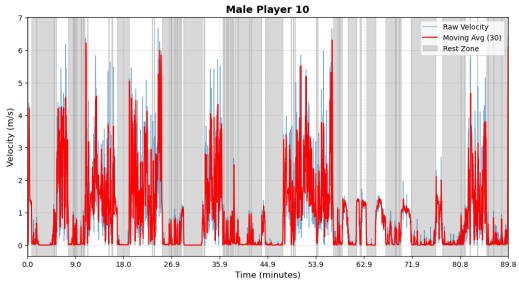


Figure A18. Male athlete 10 - Period Segmentation

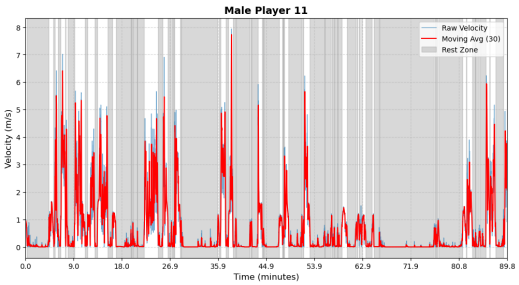


Figure A19. Male athlete 11 - Period Segmentation

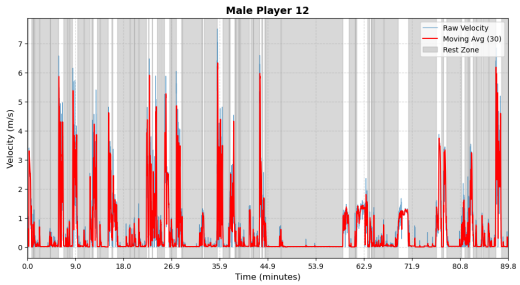


Figure A20. Male athlete 12 - Period Segmentation

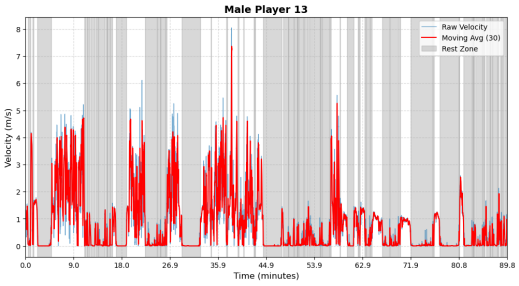


Figure A21. Male athlete 13 - Period Segmentation

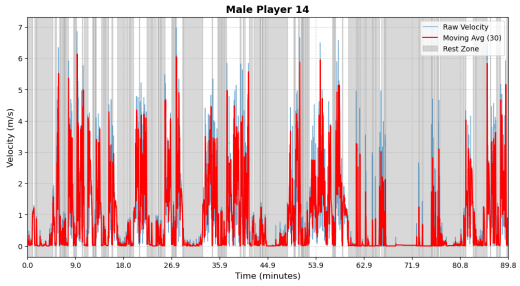


Figure A22. Male athlete 14 - Period Segmentation

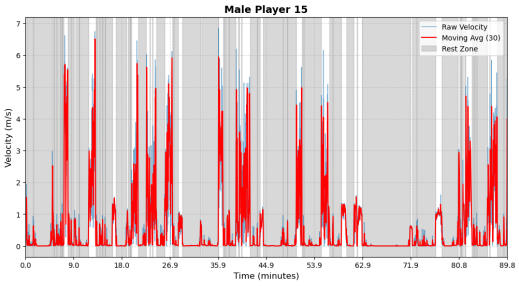


Figure A23. Male athlete 15 - Period Segmentation

Appendix B. Metabolic Power Calculation

The metabolic power model under the framework of di Prampero et al. treats forward acceleration in flat terrain as energetically equivalent to constant-speed uphill running, thus capturing both speed and acceleration-related demands [6]. Figure A24 illustrates the energetic equivalence. Panel (1) shows acceleration on flat ground, and Panel (2) represents uphill locomotion with an equivalent energy cost. The slope angle α, derived from the ratio a_f/g, underpins subsequent energy cost computations.

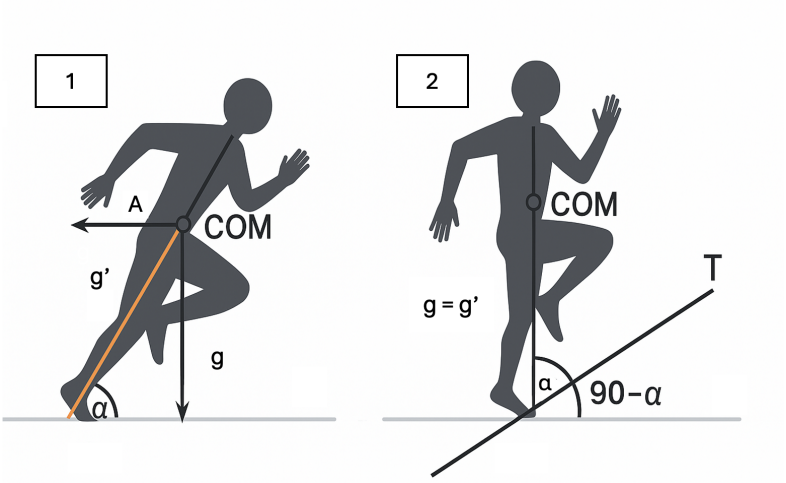


Figure A24. Energetic equivalence between forward acceleration (left) and constant-speed uphill running (right) [6]. In this Figure, COM refers to Center of Mass.

The equivalent slope (ES) is given by [6]:

$$\mathrm{ES} = \tan\left(90^\circ - \arctan\left(\frac{g}{a_f}\right)\right), \tag{A1}$$

where a_f is forward acceleration and g is gravitational acceleration.

The equivalent mass (EM), which adjusts for increased muscular effort during acceleration, is computed as [6]:

$$\mathrm{EM} = \sqrt{1 + \left(\frac{a_f}{g}\right)^2}. \tag{A2}$$

The energy cost per meter C_{sr} ($\mathrm{J{\cdot}kg^{-1}{\cdot}m^{-1}}$) is [6]:

$$C_{sr} = \left(155.4 \cdot \mathrm{ES}^5 - 30.4 \cdot \mathrm{ES}^4 - 43.3 \cdot \mathrm{ES}^3 + 46.3 \cdot \mathrm{ES}^2 + 19.5 \cdot \mathrm{ES} + 3.6\right) \cdot \mathrm{EM}. \tag{A3}$$

Finally, the metabolic power P_{met} ($\mathrm{W{\cdot}kg^{-1}}$) is calculated as follows:

$$P_{met} = C_{sr} \cdot v, \tag{A4}$$

where v is the athlete's instantaneous ground speed. This model allows for high-resolution tracking of external load.

References

1. Federation of International Touch. Basic rules of touch, 2024.
2. Johnston, R.; Watsford, M.; Pine, M.; Spurrs, A.; Murphy, A.; Pruyn, E. The Validity and reliability of 5-hz global positioning system units to measure team sport movement demands. *Journal of Strength and Conditioning Research* **2012**, *26*, 758–765. https://doi.org/10.1519/JSC.0b013e318225f161.
3. Seçkin, A.; Ateş, B.; Seçkin, M. Review on wearable technology in sports: Concepts, challenges and opportunities. *Applied Sciences* **2023**, *13*, 10399. https://doi.org/https://doi.org/10.3390/app131810399.
4. Zadeh, A.; Taylor, D.; Bertsos, M.; Tillman, T.; Nosoudi, N.; Bruce, S. Predicting sports injuries with wearable technology and data analysis. *Frontiers in Physiology* **2020**, *23*, 1023–1037. https://doi.org/https://doi.org/10.1007/s10796-020-10018-3.
5. Ferraz, A.; Duarte-Mendes, P.; Sarmento, H.; Valente-Dos-Santos, J.; Travassos, B. Tracking devices and physical performance analysis in team sports: a comprehensive framework for research—trends and future directions. *Frontiers in Sports and Active Living* **2023**, *5*, 1284086. https://doi.org/10.3389/fspor.2023.1284086.

6. di Prampero, P.E.and Fusi, S.; Sepulcri, L.; Morin, J.; Belli, A.; Antonutto, G. Sprint running: a new energetic approach. *Journal of Experimental Biology* **2005**, *208*, 2809–2816. https://doi.org/https://doi.org/10.1242/JEB.01700.
7. Impellizzeri, F.; Marcora, S.; Coutts, A. Internal and external training load: 15 years on. *International Journal of Sports Physiology and Performance* **2019**, *14*, 270–273. https://doi.org/10.1123/ijspp.2018-0935.
8. Martorelli, A.; De Lima, F.; Vieira, A.; Tufano, J.; Ernesto, C.; Boullosa, D.; Bottaro, M. The interplay between internal and external load parameters during different strength training sessions in resistance-trained men. *European Journal of Sport Science* **2020**, *21*, 16—-25. https://doi.org/https://doi.org/10.1080/17461391.2020.1725646.
9. Piedra, A.; Caparrós, T.; Vicens-Bordas, J.; Peña, J. Internal and external load control in team sports through a multivariable model. *Journal of Sport Science & Medicine* **2021**, *20*, 751–758. https://doi.org/10.52082/jssm.2021.751.
10. Beaven, R.P.; Highton, J.M.; Thorpe, M.C.; Knott, E.V.; Twist, C. Movement and physiological demands of international and regional men's touch rugby matches. *The Journal of Strength & Conditioning Research* **2014**, *28*, 3274–3279. https://doi.org/DOI:10.1519/JSC.0000000000000535.
11. Cummins, C.; Orr, R.; O'Connor, H.; West, C. Global positioning systems (GPS) and microtechnology sensors in team sports: A systematic review. *Sports Medicine* **2013**, *43*, 1025–1042. https://doi.org/10.1007/s40279-013-0069-2.
12. Hausler, J.; Halaki, M.; Orr, R. Application of Global Positioning System and Microsensor Technology in Competitive Rugby League Match-Play: A Systematic Review and Meta-analysis. *Sport Medicine* **2016**, *46*, 559–588. https://doi.org/10.1007/s40279-015-0440-6.
13. Jennings, D.; Cormack, S.; Coutts, A.; Boyd, L.; Aughey, R. The validity and reliability of GPS units for measuring distance in team sport specific running patterns. *International Journal of Sports Physiology and Performance* **2010**, *5*, 328–41. https://doi.org/10.1123/ijspp.5.3.328.
14. Varley, M.; Fairweather, I.; Aughey, R.J. Validity and reliability of GPS for measuring instantaneous velocity during acceleration, deceleration, and constant motion. *Journal of Sports Sciences* **2012**, *30*, 121–7. https://doi.org/10.1080/02640414.2011.627941.
15. Johnston, R.; Watsford, M.; Kelly, S.; Pine, M.; Spurrs, R. Validity and interunit reliability of 10 Hz and 15 Hz GPS units for assessing athlete movement demands. *Journal of Srength and Conditioning Research* **2014**, *28*, 1649–55. https://doi.org/10.1519/JSC.0000000000000323.
16. Nikolaidis, P.; Clemente, F.; van der Linden, C.; Rosemann, T.; Knechtle, B. Validity and reliability of 10-hz global positioning system to assess in-line movement and change of direction. *Frontiers in Physiology* **2018**, *9*. https://doi.org/10.3389/fphys.2018.00228.
17. Chow, C. Global positioning system activity profile in touch rugby: does training meet the match-play intensity in a two-day international test match series? *Journal of Sports Science & Medicine* **2020**, *19*, 613.
18. Vickery, W.; Harkness, A. Physical, physiological and perceptual match demands of amateur mixed gender touch players. *Journal of Sports Science & Medicine* **2017**, *16*, 589.
19. Zaragoza, J.; Prather, J.; Florez, C.; Goonan, J.; Tinnin, M.; Secrest, A.; Taylor, L.; Elsworthy, N.; Dalbo, V. 2019 Touch World Cup: An analysis of movement demands by half and gender. *International Journal of Sports Science & Coaching* **2023**, *18*, 1240–1247. https://doi.org/10.1177/17479541221100185.
20. Bourdon, P.; Cardinale, M.; Murray, A.; Gastin, P.; Kellmann, M.; Varley, M.; Gabbett, T.; Coutts, A.; Burgess, D.; Gregson, W.; et al. Monitoring athlete training loads: consensus statement. *International journal of sports physiology and performance* **2017**, *12*, S2–161. https://doi.org/10.1123/IJSPP.2017-0208.
21. Helwig, J.; Diels, J.; Röll, M.; Mahler, H.; Gollhofer, A.; Roecker, K.; Willwacher, S. Relationships between external, wearable sensor-based, and internal parameters: A systematic review. *Sensors* **2023**, *23*, 827. https://doi.org/https://doi.org/10.3390/s23020827.
22. Halson, S. Monitoring Training Load to Understand Fatigue in Athletes. *Sports Medicine* **2014**, *Suppl 2*, 139—-147. https://doi.org/10.1007/s40279-014-0253-z.
23. Windt, J.; Gabbett, T. How do training and competition workloads relate to injury? The workload—injury aetiology model. *British Journal of Sports Medicine* **2017**, *51*, 428–435. https://doi.org/doi.org/10.1136/bjsports-2016-096040.
24. Griffin, A.; Kenny, I.; Comyns, T.; Lyons, M. The association between the acute:chronic workload ratio and injury and its application in team sports: A systematic review. *Sports Medicine* **2020**, *50*, 561–580. https://doi.org/10.1007/s40279-014-0253-z.
25. Taylor, R.; Myers, T.; Sanders, D.; Ellis, M.; Akubat, I. The Relationship between Training Load Measures and Next-Day Well-Being in Rugby Union Players. *Applied Sciences* **2021**, *11*, 5926. https://doi.org/https://doi.org/10.3390/app11135926.
26. Foster, C.; Florhaug, J.; Franklin, J.; GottschallO, L.; Hrovatin, L.; Parker, S.; Doleshal, P.; Dodge, C. A new approach to monitoring exercise training. *Journal of Strength and Conditioning Research* **2001**, *15*, 109–115.
27. Hulin, B.T.; Gabbett, T.J.; Lawson, D.W.; Caputi, P.; Sampson, J.A. The acute: chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *British Journal of Sports Medicine* **2016**, *50*, 231–236. https://doi.org/10.1136/bjsports-2015-094817.
28. Akenhead, R.; Nassis, G. Training load and player monitoring in high-Level football: Current practice and perceptions. *International Journal of Sports Physiology and Performance* **2016**, *11*, 587–593. https://doi.org/10.1123/ijspp.2015-0331.

29. Hader, K.; Rumpf, M.; Hertzog, M.; Kilduff, L.; Girard, O.; Silva, J. Monitoring the athlete match response: Can external load variables predict post-match acute and residual fatigue in soccer? A systematic review with meta-analysis. *Sports Medicine* **2019**, *5*, 48. https://doi.org/10.1186/s40798-019-0219-7.
30. Hulin, B.; Gabbett, T.; Johnston, R.; Jenkins, D. High accelerations, not distances, are associated with elevated injury risk in professional rugby league. *Journal of Science and Medicine in Sport* **2016**, *19*, 332–337. https://doi.org/10.1016/j.jsams.2015.04.005.
31. Nielsen, R.O.; Bertelsen, M.L.; Ramskov, D.; Møller, M.; Hulme, A.; Theisen, D.; Finch, C.F.; Fortington, L.V.; Mansournia, M.A.; Parner, E.T. Time-to-event analysis for sports injury research part 2: time-varying outcomes. *British Journal of Sports Medicine* **2019**, *53*, 70–78. https://doi.org/doi:10.1093/pan/mpm013.
32. Kaplan, E.; Meier, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **1958**, *53*, 457–481. https://doi.org/https://doi.org/10.2307/2281868.
33. Mantel, N. Evaluation of survival data and two new rank order statistics arising in its considerations. *Cancer Chemother. Rep.* **1972**, *50*, 187–220. https://doi.org/https://www.ncbi.nlm.nih.gov/pubmed/5910392.
34. Cox, D. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **1972**, *34*, 187–202. https://doi.org/https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.
35. Gabbe, B.; Finch, C.; Bennell, K.; Wajswelner, H. Risk factors for hamstring injuries in community level Australian football. *British journal of sports medicine* **2005**, *39*, 106–110.
36. Yadav, C.P.; Sreenivas, V.; Khan, M.; Pandey, R. An overview of statistical models for recurrent events analysis: a review. *Epidemiology (Sunnyvale)* **2018**, *8*, 354. https://doi.org/10.4172/2327-4972.1000354.
37. Nielsen, R.; Bertelsen, M.; Ramskov, D.; Møller, M.; Hulme, A.; Theisen, D.; Finch, C.; Fortington, L.; Mansournia, M.; Parner, E. Time-to-event analysis for sports injury research part 2: Time-varying outcomes. *British Journal of Sports Medicine* **2019**, *53*, 70–78. https://doi.org/https://doi.org/10.1136/bjsports-2018-100000.
38. Bieszk-Stolorz, B. Relationship between coefficients in parametric survival models for exponentially distributed survival time—registered unemployment in Poland. *Econometrics* **2025**, *13*, 1. https://doi.org/https://doi.org/10.3390/econometrics13010001.
39. Mahmood, A.; Ullah, S.; Finch, C. Application of survival models in sports injury prevention research: A systematic review. *British Journal of Sports Medicine* **2014**, *48*, 630. https://doi.org/https://doi.org/10.1136/bjsports-2014-093494.190.
40. Donaldson, M.G.; Sobolev, B.; Cook, W.L.; Janssen, P.A.; Khan, K.M. Analysis of recurrent events: a systematic review of randomised controlled trials of interventions to prevent falls. *Age and Ageing* **2009**, *38*, 151–155. https://doi.org/10.1093/ageing/afn279.
41. Ozga, A.K.; Kieser, M.; Rauch, G. A systematic comparison of recurrent event models for application to composite endpoints. *BMC medical research methodology* **2018**, *18*, 1–12. https://doi.org/https://doi.org/10.1186/s12874-017-0462-x.
42. Anderson, P.; Gill, R. Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics* **1982**, *10*, 1100–1120. https://doi.org/https://www.jstor.org/stable/2240714.
43. Venturelli, M.; Schena, F.; Zanolla, L.; Bishop, D. Injury risk factors in young soccer players detected by a multivariate survival model. *Journal of science and medicine in sport* **2011**, *14*, 293–298.
44. Prentice, R.L.; Williams, B.J.; Peterson, A.V. On the regression analysis of multivariate failure time data. *Biometrika* **1981**, *68*, 373–379. https://doi.org/https://doi.org/10.2307/2335582.
45. Ullah, S.; Gabbett, T.J.; Finch, C.F. Statistical modelling for recurrent events: an application to sports injuries. *British Journal of Sports Medicine* **2014**, *48*, 1287–1293. https://doi.org/10.1136/bjsm.2003.011197.
46. Vaupel, J.W.; Manton, K.G.; Stallard, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **1979**, *16*, 439–454. https://doi.org/https://doi.org/10.2307/2061224.
47. Balan, T.A.; Putter, H. A tutorial on frailty models. *Statistical Methods in Medical Research* **2020**, *29*, 3424–3454. https://doi.org/10.1177/0962280220921889.
48. Clayton, D.G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **1978**, *65*, 141–151. https://doi.org/https://doi.org/10.2307/2335289.
49. Oakes, D. A model for association in bivariate survival data. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **1982**, *44*, 414–422. https://doi.org/https://www.jstor.org/stable/i315994.
50. Macis, A. The role of the frailty in the evaluation of injury risk factors for National Basketball Association players. *Computational Statistics* **2025**, *40*, 1985–2003. https://doi.org/https://doi.org/10.1007/s00180-024-01556-4.
51. Bowen, L.; Gross, A.; Gimpel, M.; Bruce-Low, S.; Li, F. Spikes in acute:chronic workload ratio (ACWR) associated with a 5-7 times greater injury rate in English Premier League football players: A comprehensive 3-year study. *British Journal of Sports Medicine* **2019**, *54*, 731–738. https://doi.org/doi:10.1136/bjsports-2018-099422.
52. Xu, J.; Kalbfleisch, J.; Tai, B. Statistical Analysis of Illness–Death Processes and Semicompeting Risks Data. *Biometrics* **2010**, *66*, 716–725. https://doi.org/https://www.jstor.org/stable/40962442.

53. Keiding, N.; Andersen, P.; Klein, J. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* **1997**, *66*, 215–224.
54. Bender, A.; Groll, A.; Scheipl, F. A generalized additive model approach to time-to-event analysis. *Statistical Modelling* **2018**, *18*, 299–321. https://doi.org/https://doi.org/10.1177/1471082X17748083.
55. Simon, N.; Friedman, J.H.; Hastie, T.; Tibshirani, R. Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of statistical software* **2011**, *39*, 1–13. https://doi.org/10.18637/jss.v039.i05.
56. Harper, D.; Carling, C.; Kiely, J. High-Intensity acceleration and deceleration demands in elite team sports competitive match play: A systematic review and meta-analysis of observational studies. *Sports Medicine* **2019**, *49*, 1923–1947. https://doi.org/10.1007/s40279-019-01170-1.
57. Dwyer, D.B.; Gabbett, T.J. Global positioning system data analysis: Velocity ranges and a new definition of sprinting for field sport athletes. *The Journal of Strength & Conditioning Research* **2012**, *26*, 818–824. https://doi.org/10.1519/JSC.0b013e3182276555.
58. Marutani, Y.; Konda, S.; Ogasawara, I.; Yamasaki, K.; Yokoyama, T.; Maeshima, E.; Nakata, K. Gaussian mixture modeling of acceleration-derived signal for monitoring external physical load of tennis player. *Frontiers in Physiology* **2023**, *14*, 1161182. https://doi.org/https://doi.org/10.3389/fphys.2023.1161182.
59. Richardson, A.; Welsh, A. Robust restricted maximum likelihood in mixed linear models. *Bimetrics* **1995**, *51*, 1429–1439. https://doi.org/https://doi.org/10.2307/2533273.