Predicting High-Performance Decathlon Career Best

Perry Battles[1], Tyler J. Noble[2], Robert F. Chapman[1,2]

[1] H.H. Morris Human Performance Laboratories, Indiana University Bloomington

[2] U.S.A. Track and Field

Corresponding Author

Robert Chapman

812-856-2452 | rfchapma@iu.edu

Department of Kinesiology

School of Public Health

Indiana University

1025 E 7th St, Bloomington, IN 47405

Keywords: performance modeling, track and field, elite athletes.

Word Count: 8439

Reference Count: 20

Subject Area: Environmental and exercise

New Findings

- What is the central question of this study?

What are the most important event strengths to look for in selecting and developing an elite track and field decathlete?

- What is the main finding and its importance?

Using a gamma generalized linear model, this investigation determines that first-season performances in the pole vault, javelin throw, long jump, and shot put are critical for identifying athletes with elite potential. To develop a decathlete with elite potential over a career, an approach that emphasizes improvement in pole vault and long jump is indicated to facilitate optimal career performance.

Abstract

The decathlon is a ten-event discipline in the sport of track and field, typically offered only for men at the elite level of competition (heptathlon is the complementary event for women). It is composed of the 100m, long jump (LJ), shot put (SP), high jump (HJ), 400m (day one), and the 110m hurdles, discus throw (DT), pole vault (PV), javelin throw (JT), and 1500m (day two). Using event-specific coefficients, marks are converted to scores which sum to produce an overall score. Ten events which share some underlying features, but also some disparate features, create a challenge for coaches and sport administrators in identifying potential elite talent and developing it optimally across the career. In this investigation, decathletes were profiled on the basis of keeping performances at or above 6400 points and including only those participants that competed in at least four years $(n = 642)$. Using gamma generalized linear regression models, the first-season best mark and improvement in each event were related to all-time career best (PB). Coefficients were compared to infer priorities for elite decathlete selection (using first-season best marks in each event) and development (using the delta or improvement in each event). These data show that the most critical event for identifying an athlete with a future potential elite PB is the pole vault, followed by the javelin throw, long jump, and shot put. For optimal career development, the data suggest athletes get the most return by focusing on improvement in the pole vault and long jump.

Abbreviations

- DT: Discus Throw.
- FSB: First-Season Best.
- HJ: High Jump.
- JT: Javelin Throw.
- LJ: Long Jump.

- PB: All-Time/Career Personal Best.
- PV: Pole Vault.
- SP: Shot Put.

Introduction

The decathlon is a ten-event discipline in the sport of track and field contested over two days of competition. The events consist of (in order) the 100m, long jump (LJ), shot put (SP), high jump (HJ), and 400m (all on day one), and the 110m hurdles, discus throw (DT), pole vault (PV), javelin throw (JT), and 1500m (day two). The athlete's performance mark for each individual event earns points based on a scoring table (Council, 2001 (Reprinted June 2016)), and point contributions from all ten events are added for an overall score.

Athletes don't normally become decathletes as their first-choice event in the sport of track & field. Commonly, US collegiate or international club coaches will encourage athletes with moderate (but often not elite) ability across a wide range of events to try the decathlon, as they likely could not achieve competitive national or international excellence in any one event. The challenge for coaches in identifying and training decathletes for national- or international-level success is two-fold: 1) to determine which individual events and what baseline level of ability in those events is predictive of potential future success in the decathlon, and 2) from a physiological and skill development perspective, once a potential first-year decathlete is identified, which events are the most critical to improve in to progress to an elite level in the decathlon. This latter point is important for the coach, as the physiological and technical components of the different events are not always complimentary, and determing how to logistically plan for and allocate training time for ten events can be a challenge.

The presence of multiple events with some level of shared underlying athletic abilities has focused much of the inquiry to date concerning the decathlon, with a goal of resolving these events into relevant groups and determining which of them matter the most for overall performance (Park & Zatsiorsky, 2011; Dziadek *et al.*, 2022). Intuitively, events are grouped by technique similarity and physiological demand: the sprints (100m, 110m hurdles, 400m), the jumps (LJ, HJ, PV), and the throws (SP, DT, JT), and the 1500m occupying its own category as a short endurance event. However, analysis of marks in the ten events has determined that the boundaries between them are not so clear cut. Walker and Caddigan observed that, although correlations between events were generally higher within functional groupings (sprints, jumps, throws) than without, the PV did not appear to be very associated with the other jumps and the 1500m, if grouped with the other running movements, was not correlated with the other members of its functional group (Walker & Caddigan, 2015). Additionally, the groupings may not be consistent across the entire career; Dziadek and colleagues found that, in early stage decathlon, the first principal component factor loadings indicated groupings such as speed/strength (400m, 110m hurdles, LJ, SP, DT), mixed sprint/jump/throw/endurance (JT, 1500m, 100, HJ), and another mixed jump/throw group (PV, HJ, SP). This contrasts with loadings during stage later-stage decathlon, which revealed the groupings of sprints (100m, 400m, 110m hurdles, LJ), strength/endurance (DT, SP, 1500m), jumps/throws (HJ, JT, LJ), and sprint/jump/throw/endurance (PV, 1500m, 110m hurdles, JT) (Dziadek *et al.*, 2022). These groupings not only disagree with one another, but also with the expected associations that follow from the physiological demands of the disciplines (e.g., the hurdles finding themselves in the same group as the 1500m in late stage decathlon). Although there are expected inconsistencies in such groupings due to technical factors and even training logistics, these data beg the question whether there are alternative ways to analyze the contribution of the events to overall score beyond correlation or an analysis of transfer of training via grouping.

Although determining relevant event groupings is valuable for understanding transfer of training between disciplines, it is not obvious how it might help make predictions concerning whether an athlete can or will be successful over the span of their career. It is generally agreed that success comes down to a combination of initial talent and making outsized improvements in particular events, but evidence is conflicting concerning what those events are. Additionally, the unique scoring equations for each discipline may bias evaluation of emerging athlete potential based on event area strengths (e.g., the jumps, throws, sprints, and the 1500m run).

This study aims to use existing data on elite decathletes to quantify the relationship between an athlete's career overall personal best (PB) and their first-season best marks (FSB) and improvement in each event (delta), as this can help provide objective event-specific measures for predicting emerging athlete potential.

Methods

Ethics Approval

This study used pre-existing, publicly available data, did not make use of direct measures on human or animal subjects, and therefore did not require ethics approval. This study complied with the Declaration of Helsinki.

Data Deposition

See "Data Deposition" below.

Terminology

Event – one of the ten individual events that make up the decathlon.

Mark – the performance (in time or distance) achieved by the athlete in one of the events within the decathlon.

Score – each individual event mark from the decathlon competition is converted to a score, and each of the ten scores are added to determine an overall score for the decathlon, which determines the placings within a competition.

Inclusion Criteria

Competitive performances for high-level men's decathlon from the years 2001 to 2022 were obtained from a publicly available database (World Athletics). Only overall scores above 6400 points were kept as this denoted a subjectively determined minimum value that would yield a realistic chance of becoming elite. Within this dataset, the inclusion criterion was that athletes competed on record for at least four seasons. This narrowed the sample to only those athletes that were chronically relevant from a competition standpoint, and left an overall athlete count of 642.

Data Extraction

Prior to extraction, the data were preprocessed by converting marks in the 100m, 110m hurdles, 400m, and 1500m from times to speeds such that a larger point reward in any event always corresponded to a larger mark and an improvement in any event would always correspond to a positive value. Data extraction consisted of identifying the first-season event bests (FSB) and all-time event best marks from the raw data, such that each observation in the resulting dataset consisted of an athlete profile of first season best and all-time best performances in each event. Performance deltas (delta) were then computed by subtracting the first-season best mark from the all-time best mark for each event. The highest overall meet score for the decathlete (PB) was also extracted.

Statistics

Prior to model fitting, the performance delta variables were transformed using Box-Cox scaling (Box & Cox, 1964). Values of λ were explored from -20 to 20 in increments of 0.1, and the λ among those corresponding to the highest calculated log-likelihood was selected for use in the transformation of each variable. This λ value was also rounded to the second decimal place. This resulted in λ values between 0.3 and 0.6 for each feature.

Since Box-Cox scaling does not handle negative values by default, a correction was used for the delta features to ensure there would not be negative values provided as input to the scaler. This correction factor was equal to 0 if all the values in the feature were above zero, or equal to the absolute value of the the smallest element in the feature plus 0.01 if there were elements at or below 0. This resulted in a correction of 0.01 for each delta feature.

Following this the Box-Cox transformation, all predictors were subjected to robust scaling. This additional level of scaling was selected for two reasons: (a) it has minimal assumptions compared to other modes of scaling, and (b) it is more robust to outliers compared to other common forms of scaling. Due to the nature of studying elite sport, outliers are often the most important observations in any dataset. As such, care should be taken not to distort the data or needlessly eliminate extreme observations while still allowing statistical models to provide grounds for inferential claims.

Prior to model fitting, tests for collinearity were performed on the transformed data. It was found that only two features were affected ($r \geq 0.7$): FSB_{DT} and FSB_{SP}.

Three generalized linear models were fit to the data using FSB and delta marks as covariates and PB as the outcome variable. The first model was fit with a log link, the second with the inverse link (the canonical linking function for gamma generalized linear models (Dunn *et al.*, 2018, Table 6.4)), and the third model was fit using the identity link function. Visual inspection of residuals was used to assess model appropriateness and to verify model assumptions.

Subsequent to model fitting, influential outliers were identified using a panel of tests. These included the DFBETAS and DFFITS for each model variable, the covariance ratio, and Cook's Distance, as described in Belsley *et al.* (1980), Cook & Weisberg (1982).

Model coefficients for FSB and delta values were rank-ordered by magnitude to compare and contrast their effects on PB. Subsequent to fitting, Spearman's ρ was used to compare the ordering of event coefficients to assess agreement between the models. For all statistical tests conducted in this investigation, α was set to 0.05.

Importantly, this is an exploratory investigation by nature. Although the comparison of many coefficients across three separate models will certainly wildly inflate familywise error, it is in the nature of exploratory investigation to often forego a hypothesis and compare many factors which may influence the outcome of interest (Sullivan & Feinn, 2021). This is especially true in light of the fact that investigations into the decathlon have either found inconsistent or conflicting results within themselves (Dziadek *et al.*, 2022) or between one another (Van Damme *et al.*, 2002; e.g., Kenny *et al.*, 2005). As such, this paper performs pairwise comparisons to more comprehensively explore the relationships between model predictors.

To compare model coefficients, a bootstrapping technique was applied to the difference in estimates of two coefficients. For each of 1000 bootstrap replicates of the data (sans outliers), a model with the appropriate link function was fit to the bootstrapped sample, relating all predictors to the outcome. The difference for the pair of coefficient estimates in question was extracted and used in the construction of a 95% confidence interval. If the confidence interval of the difference between the two coefficient estimates crossed zero, there was no significant difference at $\alpha = 0.05$. If the lower bound of the confidence interval was above 0, this constituted a significant difference in which the first coefficient in the pair was greater than the second. If the upper bound of the confidence interval was below 0, this indicated that the first coefficient was significantly lesser than the second. This process was repeated with 1000 boostrap replicates for each coefficient comparison. The results are displayed in figs. 2 to 4 and figs. 6 to 8. Bias-corrected accelerated confidence intervals were used for the comparison of model coefficients.

One important consequence of this bootstrapping approach to coefficient comparison is that a row-vs.-column comparison can have a different outcome than a complementary column-vs.-row comparison of coefficients, especially when the confidence interval for the comparison in question is very close to zero (e.g., in the inverse link model delta coefficients, the PV was significantly more influential than the 110m hurdles on one axis but not on the other). This affected interpretation of differences between the HJ/400m, HJ/JT, 1500m/400m among the FSB coefficients, and comparisons of the PV/110m hurdles and 1500m/JT among the delta coefficients.

Importantly, while it is true that a $100 * \alpha$-level confidence interval that does not contain 0 corresponds to a significant difference at $p \leq \alpha$ and has an accompanying p value, the volume of comparisons performed in the present study means that there would be 270 such p values reported. As such, these have not been presented in the text, but indicated in figures using color- and hatch-coding to signify significant differences.

Results

Model Appropriateness

Standardized Deviance Residuals

The plot of the standardized deviance residuals against the log of the fitted values did not appear to display any strong discernable trends for either the log link or identity link models. The plot for the inverse link model, however, appeared to display a trend.

Linear Predictor vs. Working Residuals

A plot of the linear predictor against the working residuals demonstrated high model appropriateness with respect to the data for all three models.

Partial Residual Plots

Partial residual plots further confirmed model appropriateness, with errors for all features following the ideal relationship. Those that deviated the most from this relationship were the delta values for the 110m hurdles, 400m, shot put, and 1500m.

Model Coefficients

Once models were refit after outlier removal, coefficients for all first-year event marks and career delta were significant at $p \leq 0.0001$ with the exception of $\text{delta}_{100\text{m}}$ in all three models, which was significantly related to the outcome variable in the log model at $p = 0.000$, at $p = 0.001$ in the inverse link model, and at $p = 0.000$ in the identity link model.

First-Season Bests

Figure 1 shows the FSB coefficients, sorted by absolute magnitude, for each model. The models disagree slightly about the positions of the HJ and the SP and the 400m and the 100m. The pole vault is ranked highest across all three models, followed by either the long jump or the javelin throw depending on the model. There was a high degree of correspondence between the log link model and the inverse link model (Spearman's $\rho = -0.995, p \leq 0.05$) and between the identity link model and the inverse link model (Spearman's $\rho = -0.982, p \leq 0.05$). The same was true of the log and identity link models (Spearman's $\rho = 0.989, p \leq 0.05$).

Figures 2 to 4 show which coefficients are significantly different from the others at $p \leq 0.05$. For example, the coefficient for the PV was significantly greater ($p \leq 0.05$) (or smaller in the inverse model given the nature of the linking function; subsequent indication of this difference will henceforth be omitted) than all others across the three models ($p \leq 0.05$). Additionally, the coefficients for the LJ, PV, SP, and JT were all significantly greater ($p \leq 0.05$) than the coefficients for the 100m sprint and 400m sprint in all three models. The coefficient the JT was significantly greater ($p \leq 0.05$) than that of the 110m hurdles in the log and identity link models. Additionally, considering the 100m sprint alone, the LJ, HJ, PV, SP, JT, and 1500m all had coefficients that were significantly greater than that of the 100m ($p \leq 0.05$) across all three models. The DT is also included in this group if only the log and inverse link models are considered.

Several events were not significantly different from one another; these consist chiefly of events that share the same group (e.g., 100m, 400m, or DT, JT). For further examples, see figs. 2 to 4.

Deltas

Figure 5 shows the delta coefficients, sorted by absolute magnitude, for each of the models.

There were minimal differences between the models in terms of coefficient comparisons. The only difference between the results for the log link and identity link models was in a comparison between the 1500m and the JT, which indicated that the JT was significantly less influential than the 1500m in the identity link model (fig. 8). This difference also held between the inverse link model and the two other models, in addition to the fact that the LJ was significantly more influential than the 400m in the inverse link model (fig. 7). There was also a mistmatched comparison in the inverse link model in which the PV was significantly more influential than the 110m hurdles along one axis but not the other (fig. 7).

Apart from these differences, the patterns found in the delta coefficients are generally similar those of the FSB coefficients. The pole vault had a significantly greater coefficient than all of the sprints (with the exception of the 110m hurdles in the inverse link model on one axis of comparison) and all of the throws ($p \leq 0.05$), but not the other jumps or the 1500m (figs. 6 to 8). The LJ had a significantly greater coefficient than the 100m, DT, and JT at $p \leq 0.05$ across all three models. Finally, the 100m had a significantly smaller ($p \leq 0.05$) coefficient than the 110m hurdles, LJ, HJ, PV, SP, and 1500m across all three models.

Data Deposition

...

Discussion

The decathlon coach Zdeněk Váňa concluded his 2003 presentation on the success of his team, which included some of the most elite decathletes of all time, as follows:

> I am just an ordinary coach, I started coaching youths and then the sprints and hurdles. I was very lucky and fortunate with the choice of athletes, who wanted to work hard so we made it to world-class performances and breaking the mythical 9000 point barrier in the decathlon, the royal event of track and field.
>
> (Vana, 2003 p.30).

Váňa emphasizes his good fortune in selecting excellent athletes early on and attributes their subsequent success to this. Although there is certainly more to the achievements of his cohort than simply choosing the best athletes early on, there is no doubt that identifying excellent talent played a substantial role in their achievement. Váňa credits this to his good fortune; however, for the coach and sport administrator looking to identify and develop elite talent in this complex discipline, chance cannot be relied upon to facilitate success. Instead, identifying the factors that predict long-term career success and quantifying them can help optimize talent identification and development.

In the ten-part track and field decathlon, the presence of multiple events with shared underlying characteristics makes the prediction of future performance a multidimensional problem. Additionally, since the decathlon is an event that draws talent from many other disciplines, it is precisely the ability to identify outlier potential based on individual event performances that is critical. Once talent has been selected, the next question that follows is how to invest time and effort to maximize the odds of producing a champion. As these questions have not yet been thoroughly addressed with predictive mathematical modeling, this paper uses a tool that is novel in this research niche to predict long-term career achievement in the track and field decathlon. This investigation builds upon past research, which has established associations among the events, attempted to identify trade-offs among them, and profiled elite performers. The data presented here have determined that the pole vault and, to a much lesser extent, the javelin throw are the two factors that distinguish decathletes with elite career potential from their peers, with the long jump and shot put following as secondary priorities. These data also indicate that improvement in the pole vault and the long jump set the elite decathlete apart from their peers.

Previous Research in the Decathlon

Cox and Dunn performed a multifaceted analysis of the decathlon events with the goal of identifying groupings among them and determining whether the decathlon was truly fair (in the sense that it rewarded excellence in each of the events equally) (Cox & Dunn, 2002). Using data from 1991, 1993, 1995, 1997, and 1999, they first performed cluster analysis, obtaining the following event groupings for the combined data across all years:

- 100, 400m, LJ, 110m hurdles.
- SP, DT, JT, PV.
- HJ, 1500m.

Importantly, the form of cluster analysis used by Cox and Dunn can produce different results depending upon the level at which the clusters are evaluated. For example, two clusters could just as easily have been produced, which would have resulted in (a) 100m, 400m, LJ, 110m hurdles, and (b) SP, DT, JT, PV, HJ, 1500m. However, since these groupings simply reflect associations at different levels of similarity, both are valid. The groupings they found point to associations between the events which group the sprints together (especially when considering the long jump as a sprint), and does the same for the throwing events.

In addition to their cluster analysis, Cox and Dunn investigated potential advantage for particular events due to higher variance among their scores. Reasoning that greater variance might be indicative that higher scores are awarded to a specialist in that event, they used Bartlett's test for homogeneity of variance and the Kruskal-Wallis test to compare the events. Bartlett's test revealed that the the three events with the lowest variance were track events (400m, 100m, 110m hurdles), while the field events, particularly the PV, had higher variances. On the basis of the Kruskal-Wallis test, they determined that the median scores for the ten events were not equal. However, an event is fair as long as it rewards all athletes equally (Cox & Dunn, 2002); it must not necessarily have a median score equal to that of the other disciplines to do so. It could be that all athletes acquire more points from an event without that event rewarding them unequally.

To further assess the fairness of the decathlon, they utilized Kendall's coefficient of concordance to quantify agreement between the positions of each athlete across the events. In four of the five years the value of the coefficient was significant, indicating that the placings for the athletes were in agreement across the ten events for those years. Using Spearman's ρ, they further measured the relationship between placings in the 10 events, determining that four of the five highest values of ρ were observed among the field events.

Finally, Cox and Dunn used a cumulative sum to assess event fairness. By computing the proportion of overall score due to each event and subtracting the mean, profiles for each event can be established showing the pattern from the highest-ranked decathletes to the lowest. For example, if athlete i received a very large proportion of their overall score from an event j, such as 0.15 of their total score, and the average athlete only received 0.09 of their total points from that event, this would be reflected as a value of 0.06 for those values of i, j (mathematically, $y_{ij} - \bar{y}_j$). Conversely, this value would be negative if an athlete received a lower proportion of their score from a particular event relative to the average. If many high-ranking athletes derive a large proportion of their total scores from the 100m, then the 100m should show a steeply-rising cumulative sum as positive values accumulate among the higher placings. This would then be expected to return back to 0 as the lower-ranking athletes would have negative values with respect to the mean in this regard. On the basis of this analysis, Cox and Dunn found that, although there are no individual events that consistently demonstrate the favorable "mountain" profile or the unfavorable "valley" profile consistently year to year, the favorable shapes all occur among the field events, while the unfavorable shapes all occur among the track events.

Overall, the results of Cox and Dunn indicate that the events cluster on the basis of their technical and physiological features, and that the field events are favored over the track events. This agrees with the data presented here, which point to the prevalence of the PV, JT, LJ, and SP in long-term decathlon achievement. It is possible that the underlying similarities mentioned above give rise to the prominence of these events via transfer of training. If improving in one of these disciplines increases performance in another in some meaningful way, it seems logical that these disciplines would become the most common areas of emphasis for successful athletes. For example, three of the events that are most important according to the FSB coefficients feature a run-up (PV, JT, LJ). The javelin throw, in addition, is a throwing event, and shares technical and physiological features with the SP and DT. These other throwing events do not feature a run-up and would thus seem to have less potential for transfer of training than the javelin would to the other events. In other words, by training the javelin, one could not only improve in that discipline, but also work on the run-up (shared with the PV, LJ, and HJ) and the rotational mechanics of throwing (shared with the SP and DT). This also agrees with Zdeněk Váňa's contention that the most important attributes for success in the decathlon are speed (required for the run-up), then strength (required for the throws), and then technique (Vana, 2003). However, if the run-up is a prominent shared feature among the highest-priority events, the relatively low stature of the sprints becomes an inconsistency. It is possible that the pure running events are too narrow in their training emphasis and thus become less important compared to those events which can emphasize multiple facets of performance at once when trained. It is also not necessarily the run-up that is the key shared feature, but perhaps some other component. It may also be the case that decathletes are generally more proficient in the sprints but that abilities across the other events vary more widely, making those other disciplines more deciding. Although it is not truly possible to infer on the basis of the current data what these underlying features are, there is evidence that they exist (Cox & Dunn, 2002; Park & Zatsiorsky, 2011).

Park and Zatsiorsky also utilized principal component analysis (Park & Zatsiorsky, 2011), revisiting the 1962 analysis by Zatsiorsky and Godik (Zatsiorsky & Godik, 1962). Park and Zatsiorsky concluded on the basis of their analysis that approximately 70% of decathlon performance can be attributed to three latent factors: sprinting performance (100m, 400m, 110m hurdles, LJ), throwing/jumping ability (JT, SP, DT, HJ, PV), and endurance (1500m). These results agree to some extent with the data presented here, since the first factor contains the LJ, which was an important event among both the FSB and delta features. However, the PV, which was by far the most valuable event in the present analysis, only occurred in the second principal component. This component, however, did also contain the JT, which would be considered a training priority in the present analysis. Hence, there is moderate agreement between the results of Park and Zatsiorsky and this investigation if one considers the amount of total variance that at event contributes to be indicative of its importance.

These factor loadings don't provide any indication of how much each event contributes to the overall score, but they do indicate where there is potential for transfer of training. Importantly, several of the most important FSB and delta events (SP, PV, JT) were loaded on the same principle component (the second) in the analysis performed by Park and Zatsiorsky. This indicates that there may be shared underlying features among these events that contribute to their importance. The factor loadings they found also agree with the training priorities described by Zdeněk Váňa (Vana, 2003) insofar as a decathlete that is "ahead of the curve" according to those priorities might exhibit high performance in the most technical events and greater throwing ability.

Pavlović and Idrizović used factor analysis to look at associations between the event performances of ten world record holders and inferred from their results that there were three essential elite decathlete archetypes: the "jumper-thrower-runner" (high factor loadings of PV, HJ, DT, 400m, SP), "runner-sprinter" (100m, 110m hurdles), and "jumper-thrower" (LJ, 1500m) (Pavlović & Idrizović, 2017). There is a degree of choice concerning how many factors should be considered, resulting in that number of elite decathlete profiles. However, a choice of either three or four archetypes is sensible given the three to four broad categories of events (running, jumping, throwing for 3 types; sprinting, jumping, throwing, endurance if four are considered). The results of Pavlović and Idrizović suggest the presence of distinct sub-types of elite decathletes. As before, one factor contains the pole vault and a secondary priority identified by the present analysis (SP). This could be interpreted as support for the idea that the shared features among these events are related in some way to their importance for long-term success in the decathlon. Pavlović and Idrizović conclude that the areas of greatest strength for elite decathletes are the long jump, 110m hurdles, 100m, and the pole vault, with the weaker areas being the high jump, javelin throw, shot put, discus throw, and 1500m. It is difficult to draw comparisons between the work of Pavlović and Idrizović and this investigation due to the fact that they address different questions using distinct samples and predictors. Their data concern the marks of Olympic champions at the peak of their careers performing at the Olympic Games, while the predictors included here consist only of first-season achievement and improvement across the career for decathletes of a much wider range of ability levels. On the basis of these facts alone, many speculative disparities in results could be explained.

Pavlović and colleagues (Pavlović *et al.*, 2020) then addressed an adjacent topic by comparing the best decathlon results of several elite athletes to their best results in each individual event. They concluded that decathletes tended to significantly outperform their decathlon bests with their individual bests in each event, particularly in the HJ, 110m hurdles, DT, PV, and 1500m. These results are interesting as several of these are events that are anecdotally considered more technical (PV, HJ, 110m hurdles); furthermore, four of those five occur on the second day of competition (110m hurdles, DT, PV, and 1500m). The fact that elite decathletes tend to outperform their decathlon bests by a greater margin in these events may indicate greater underlying ability, or could simply mean that they are more fatigued from the full day of competition already behind them. These events also contain the PV, the greatest priority in the present analysis. The data of Pavlović and colleagues suggest the possibility that elite decathletes tend to perform well in the events that will be affected by the fatigue from the first day, as indicated by the high discrepancy in performance or "performance reserve" that they exhibit.

Finally, Dziadek and colleagues (Dziadek *et al.*, 2022) expanded upon these analyses by introducing career stage. Their analysis was arguably influenced by differing samples sizes between stages, but revealed the potential for varying associations between events at different levels of decathlete advancement. Their factor analysis identified that the first stage was characterized by the following:

- Speed/strength component (400m, 110m hurdles, LJ, SP, DT)
- Mixed component (JT, 1500m, 100m, HJ)
- Jump/throw component (PV, HJ, SP)

At the subsequent stage ("pursuit of athletic excellence") the loadings indicated the following components:

- Speed component (110m hurdles, 400m, LJ)
- Endurance/strength/speed component (1500m, DT, 100m)
- Speed/strength component (SP, 100m, JT, HJ, DT, 400m)
- Jump/endurance/speed component (PV, LJ, 1500m, 400m, 110m hurdles)

For the "athletic excellence" stage, the loadings outlined the following components:

- Speed component (100m, 400m, 110m hurdles, LJ)
- Strength/endurance component (DT, SP, 1500m)
- Jumps/throws component (HJ, JT, LJ)
- Jumps/endurance/speed/strength component (PV, 1500m, 110m hurdles, JT)

The final career stage, characterized by declining performance, displayed factor loadings as follows:

- Speed/jump component (100m, 400m, PV, 110m hurdles)
- Endurance/strength component (1500m, DT, SP)
- Jump/throw component (LJ, JT)

The shifting groupings from stage to stage might be interpreted as reflecting changing associations of events at each career stage. Dziadek and colleagues conclude that speed and strength are the two biggest pillars of decathlon performance, a notion which is supported by the genetic data (Ben-Zaken *et al.*, 2022; Remmel *et al.*, 2023) and by the opinion of at least one elite decathlon coach (Vana, 2003). Our results partially agree with this notion, considering that the LJ and JT were among the events whose coefficients were significantly greater ($p \leq 0.05$) than those of some others. Our data also strongly point to the prominence of the pole vault, which contradicts Dziadek's findings insofar as the PV was typically part of later principal components that only explained a relatively small proportion of the total variance in their dataset.

The same can also be said to some extent of the JT.

First-Season Bests

The results of this study indicate that the PV is by far the best indicator of long-term achievement in the decathlon. FSB values that are decidedly less important than the pole vault but still valuable are the JT, LJ, and SP. There is also evidence that the 100m and 400m sprints are the least impactful indicators of career achievement compared to the other events (figs. 2 to 4).

Superficially, this contradicts certain established track and field training principles (Vana, 2003). However, it is possible that elite and non-elite professional decathletes are similarly advanced in the sprints and that it is the mastery of the other elements that Váňa mentions (strength, followed by technique) that sets champions apart. Mastery of the pole vault could be an indication that the decathlete has already become fast and strong and is now focusing on refining technique, particularly in the most technically challenging events. In this model, such disciplines (e.g., pole vault) are a proxy for overall decathlon advancement. According to the interpretation above, this could indicate that high performances in the pole vault are most predictive of high long-term career achievement because they indicate emphasis on technical practice that can only occur once strength and speed have been built and technique can solidify. The throws would be a natural second priority here since high performances in them would indicate that speed has already been established, permitting emphasis on throwing. Conversely, poor performances in the throws might indicate that speed – the earliest stage of development – is still in need of relatively high training focus, placing the decathlete behind relative to the competition. This would explain why the pole vault – and, secondarily, the throws – are more strongly influential on predictions of career PB than the sprints. A simpler explanation might be that higher first-season performance in the pole vault is simply an indication of prior experience in the event, and that this is advantageous in its own right since it is anecdotally one of the most difficult to master.

Finally, it bears mentioning that there were several events which were not significantly different from one another (e.g., DT and HJ, 400m and 100m in figs. 2 to 4). Hence, although there are certain outstanding priorities (such as the PV) and non-priorities (such as the 100m), several event pairings exist for which definitive comparisons cannot be made on the basis of these data. This lends support to the notion that the decathlon is an event which effectively rewards performance in each of the ten disciplines relatively equally (somewhat in contrast to the assertions of Westera (Westera, 2006)). Hence, while these data indicate that high first-season marks in the PV, JT, LJ, and SP are important for greater long-term career achievement, they also point to the notion that several of the events are likely similarly important.

Improvement across the Career

The priorities for the development of elite talent indicated by our data are somewhat narrower than those for early career selection. The pole vault and long jump are still prominent, but the SP is only significantly more influential than the 100m; furthermore, the JT does not have a significantly more infuential coefficient than any other event in any model. In agreement with the FSB results, the delta_{100m} is significantly less influential than the 110m hurdles, LJ, HJ, PV, SP, and 1500m.

Additionally, much like the first-season best coefficients, several of the delta coefficients are not significantly different from one another (e.g., DT and JT, 100m and 400m). Hence, there are not clear-cut statistical grounds to prioritize one over the other; this decision must be made based on the athlete's individual strengths, weaknesses, and training context.

Application of these Findings

These results may be interpreted as a function of the career stage of the athletes in question. As noted by decathlon coach Zdeněk Váňa, there is a sequence of priorities for developing decathletes that begins with speed, followed by strength, and ending with technique (Vana, 2003). Since early-stage professional decathletes often come directly from the university athletic setting and may not have competed in the decathlon prior to university, it is possible that they are at a level of advancement in which the best decathletes will be more experienced or advancing more quickly and, hence, already focusing on technique, while the less experienced or more slowly advancing decathletes will still be focusing more on speed and strength. This would explain why the pole vault, a very technical event, is most important both for first-season prediction and development of a high career PB. The secondary prominence of the SP and JT confirms this as strength is developed before technique in this model. Hence, performance in highly technical events would be indicative of an advanced or well-developed athlete, followed by performance in the throws. Conversely, athletes that are still concentrating primarily on the sprints might be "behind" for the post-university career stage from this standpoint.

The data of Cox and Dunn also support this notion. Three events whose scores have the lowest standard deviations (400m, 100m, 110m hurdles) are in the top 4 for largest median scores. High variance can be taken as an indication that an event is rewarded unfairly for outsized performances because of a favorable scoring equation, but could also indicate that these are events in which the abilities of athletes vary more widely than in others. This suggests that some athletes are much more proficient than others and points to potential asymmetries in the selection and development of the decathlete. It is noteworthy that Zdeněk Váňa started off coaching youths in sprinting events (Vana, 2003), that he suggests a sequential model of decathlete development that begins with speed as the foundation, and that decathletes are anecdotally often recruited from the ranks of sprinters (Wang, 2017). If these assumptions are made, low variance and high median scores in the sprinting events are a logical consequence since most decathletes are proficient in them – they would logically derive a higher score from those events than they do from the throwing and jumping events since they are trained in and recruited from the sprinting events. Furthermore, this foundation in the sprints would mean that there would be few decathletes that lack such proficiency. The other events, however, from which decathletes are not recruited and which are further along in the sequential athlete development model, would exhibit higher variance since some decathletes would be much better than those that are less advanced and since they may not be as disposed towards high performances in those events due to their selection from a pool of sprinters. The data presented here also support this, since coefficients for FSB performances in the LJ, PV, SP, and the JT are significantly ($p \leq 0.05$) greater than those for the 100m and the 400m, for example.

Alternatively, this could simply be an indication that the field events are favored, as determined by Cox and Dunn (Cox & Dunn, 2002). All of the events given priority by the present analysis (pole vault, long jump, javelin throw, shot put) are field events. Contrastingly, the track events appear not to be favored. This could be due to the nature of the attempt structure of the field events, which allow for the best mark among several to be chosen in place of a single performance. Alternatively, although somewhat less likely, there could be mathematical grounds for this apparent imbalance (Westera, 2006). The data presented here provide some statistical indication that improving in one subset of the events is likely much better than improving in another, and that certain initial event strengths will lead to higher long-term achievement compared to others. This could be interpreted as support for Westera's argument that the decathlon may be mathematically unfair (Westera, 2006).

There is also a possibility that athletes that tend to excel in certain events are preferentially selected for participation because they are subjectively judged to have greater potential. Contrasting Westera's more mathematically deterministic notions of decathlon achievement, this view would suggest that judgements about an athlete's potential made on the basis of their performance in certain events may play a prominent role in determining which athletes continue to compete in the decathlon and achieve high career personal bests. This could be due to increased encouragement and social or financial support, or to another socially-mediated factor. Although speculative, it is important to retain the notion that long-term achievement in the decathlon may not be merely a function of skewed scoring coefficients, as Westera suggests (Westera, 2006), but could also be a function of other influences that don't strongly depend upon the mechanics of scoring. If this is the case, the models presented here would reflect not only the decathlon scoring coefficients, but also the judgements of coaches, peers, and sport administrators.

There is also the confounding factor of the existing training priorities in the decathlon. If coaches, particularly the coaches of elite athletes, have chosen to prioritize certain events, this would suggest that these events would naturally become associated with elite decathlete development. In other words, it may be that elite coaches consistently choose similar training priorities for their athletes, and that this shapes the profile of the elite decathlete (as treated in (Pavlović & Idrizović, 2017)) artificially.

Directions for Future Research

Importantly, there is a major unanswered question concerning a key component in the strategy of elite decathlete development; namely, the difficulty of improving in each event. For example, it may be that there is tremendous point yield to be obtained from improving only a very small amount in a given event, but that even such a minute improvement is a nearly insurmountable task for the typical high-level decathlete. In this situation, it would be beneficial to weight the yield of an improvement in a mark by the feasibility of that improvement; this is something which this investigation has not attempted and which would be highly beneficial in developing future elite decathletes.

Another question that is left open by the present research is why the most prominent events have risen to the top. It could be, as suggested by the data of Cox and Dunn, that there is larger variation in the scores of those events, which would indicate a greater difference between a high and low performer. It could also be the case, as Westera contends, that these differences are driven by the scoring equations rather than by physiological or technical features. Contrarily, it could be the case that transfer of training is responsible. Answering this question would permit more informed decisions in training the elite decathlete.

Finally, it is not currently known how much of the associations between high marks in certain events and all-time career PB are due to choices in training rather than physiological, technical, or biomechanical factors. Answering this question would help to frame the importance of more innate biological features compared to training choices.

Conclusion

Taken together, these data outline key priorities and non-priorities for success in the professional decathlon. When selecting decathletes, the pole vault is likely the single most important indicator for choosing future champions, followed somewhat distantly by the javelin throw. Additionally, the long jump and shot put are generally more important than 100m and 400m and can be considered secondary priorities. In fact, the 100m, for example, is less important than the 110m hurdles (in one model), LJ, HJ, PV, SP, DT (in two out of three models), JT, and 1500m in terms of elite decathlete selection.

Once an athlete has been selected, the pathway for development should emphasize the pole vault and long jump.

Overall, the pole vault and long jump are indicated as the highest priorities for decathlon performance.

References

Belsley DA, Kuh E & Welsch RE (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

Ben-Zaken S, Meckel Y, Remmel L, Nemet D, Jürimäe J & Eliakim A (2022). The prevalence of IGF-i axis genetic polymorphisms among decathlon athletes. *Growth Hormone & IGF Research* **64,** 101468.

Box GE & Cox DR (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **26,** 211–243.

Cook RD & Weisberg S (1982). Residuals and influence in regression.

Council I (2001 (Reprinted June 2016)). *Scoring tables of athletics*. IAAF Athletics.

Cox TF & Dunn RT (2002). An analysis of decathlon data. *Journal of the Royal Statistical Society Series D: The Statistician* **51,** 179–187.

Dunn PK, Smyth GK, et al. (2018). *Generalized linear models with examples in r*. Springer.

Dziadek B, Iskra J, Mendyka W & Przednowek K (2022). Principal component analysis in the study of the structure of decathlon at different stages of sports career. *Polish Journal of Sport and Tourism* **29,** 21–28.

Kenny IC, Sprevak D, Sharp C & Boreham C (2005). Determinants of success in the olympic decathlon: Some statistical evidence. *Journal of Quantitative Analysis in Sports*.

Park J & Zatsiorsky VM (2011). Multivariate statistical analysis of decathlon performance results in olympic athletes (1988-2008). *World Academy of Science, Engineering and Technology* **5,** 985–988.

Pavlović R & Idrizović K (2017). Factor analysis of world record holders in athletic decathlon. *Sport Science* **10,** 109–116.

Pavlović R, Vrcić M & Petrović B (2020). Athletic decathlon: Are there differences between the results of decathlon recordholders and their best personal results.

Journal of Physical Education Research **7,** 18–26.

Remmel L, Ben-Zaken S, Meckel Y, Nemet D, Eliakim A & Jürimäe J (2023). The genetic basis of decathlon performance: An exploratory study. *The Journal of Strength & Conditioning Research* **37,** 1660–1666.

Sullivan GM & Feinn RS (2021). Facts and fictions about handling multiple comparisons. *Journal of Graduate Medical Education* **13,** 457–460.

Van Damme R, Wilson RS, Vanhooydonck B & Aerts P (2002). Performance constraints in decathletes. *Nature* **415,** 755–756.

Vana Z (2003). The training of the best decathletes. *New Studies in Athletics* **18,** 15–30.

Walker JA & Caddigan SP (2015). Performance trade-offs and individual quality in decathletes. *Journal of Experimental Biology* **218,** 3647–3657.

Wang Z (2017). Men's decathlon high level athlete's performance characteristics. In *2016 national convention on sports science of china*, p. 01032. EDP Sciences.

Westera W (2006). Decathlon, towards a balanced and sustainable performance assessment method. *New Studies in Athletics* **21,** 39–51.

Zatsiorsky V & Godik M (1962). Mathematics and decathlon. *Track and Field* **10,** 28–30.

Additional Information

...

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Author Contributions

Contributions/roles are as follows:

- Robert F. Chapman: conceptualization, methodology, project administration, writing (review and editing).
- Tyler J. Noble: conceptualization, data curation.
- Perry S. Battles: conceptualization, formal analysis, methodology, software, verification, visualization, writing (original draft).

Artificial Intelligence Generated Content

None of the content of this article or the results outlined herein were generated by a large language model or other form of generative artificial intelligence.

Author Pronouns

Perry Battles (he/him/his).

Tyler J. Noble (he/him/his).

Robert Chapman (he/him/his).

Funding

No funding was received or provided by the authors for the completion of this manuscript.

Acknowledgements

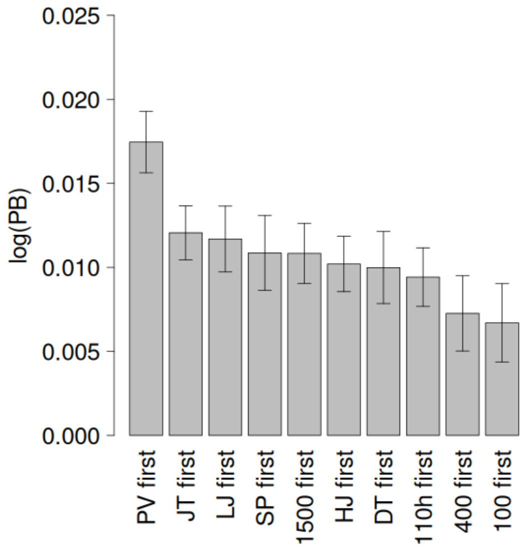
. . .

Tables

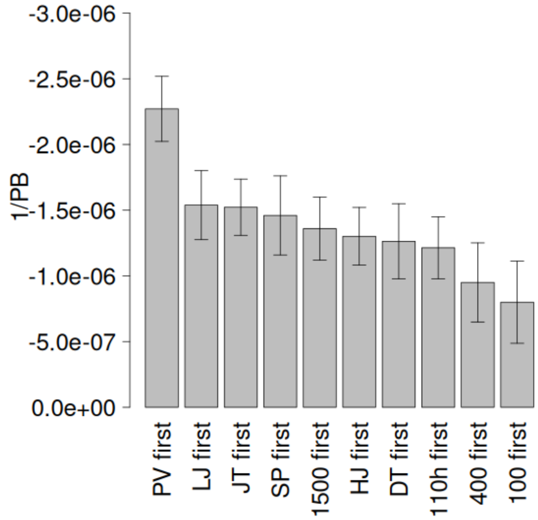
. . .

Figures and Legends

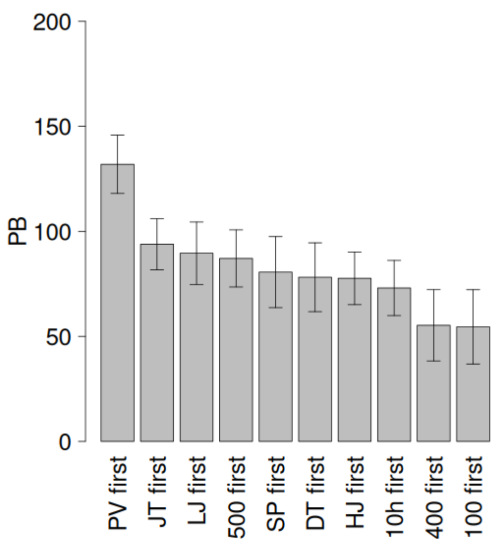
Figures and legends to follow.



(a) Log Link



(b) Inverse Link



(c) Identity Link

Figure 1: Absolute Magnitudes of FSB coefficients, sorted by absolute magnitude from highest to lowest. Bars indicate confidence interval.

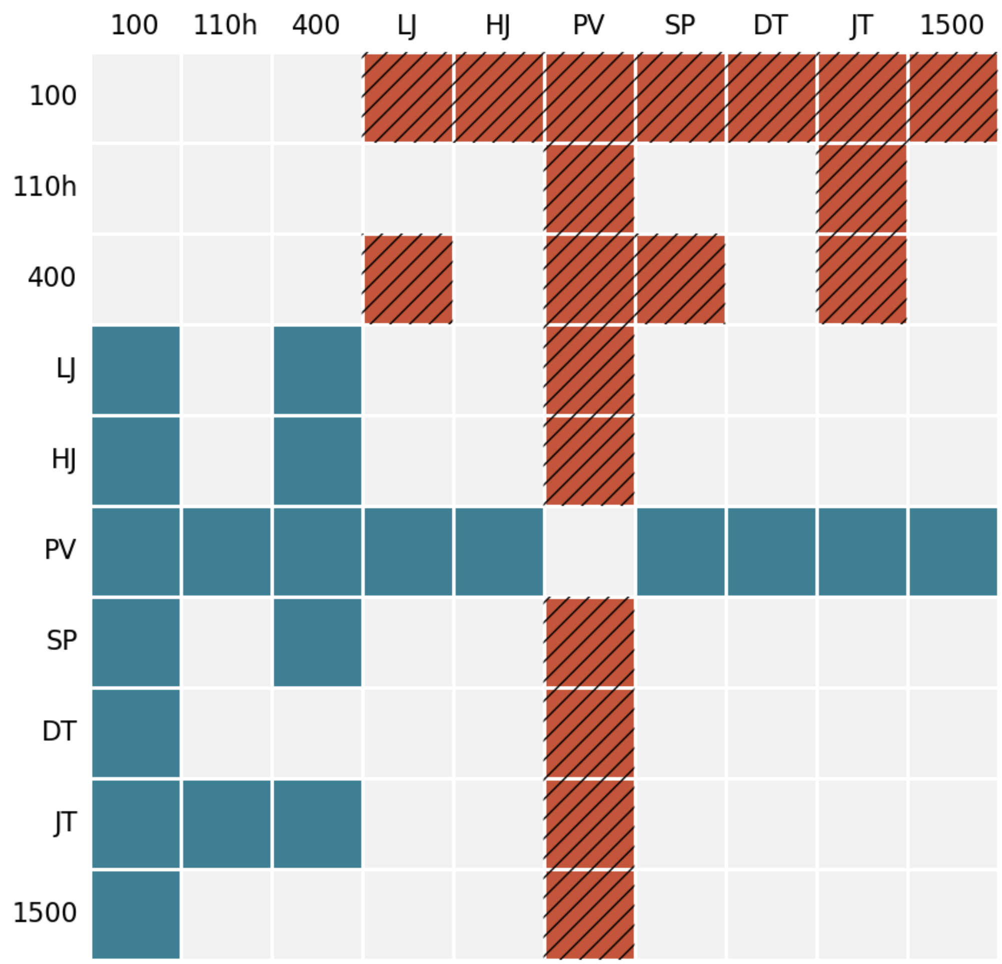


Figure 2: Matrix diagram showing significant differences between FSB coefficients, log link . A hatched orange square indicates that the entirety of the bootstrapped confidence interval of the difference between the row coefficient and column coefficient is below zero, indicating the column coefficient is greater. A teal square indicates that the entire confidence interval is above 0, meaning that the row coefficient is greater. A grey square indicates that this interval contains 0, which means that neither one is significantly greater than the other.

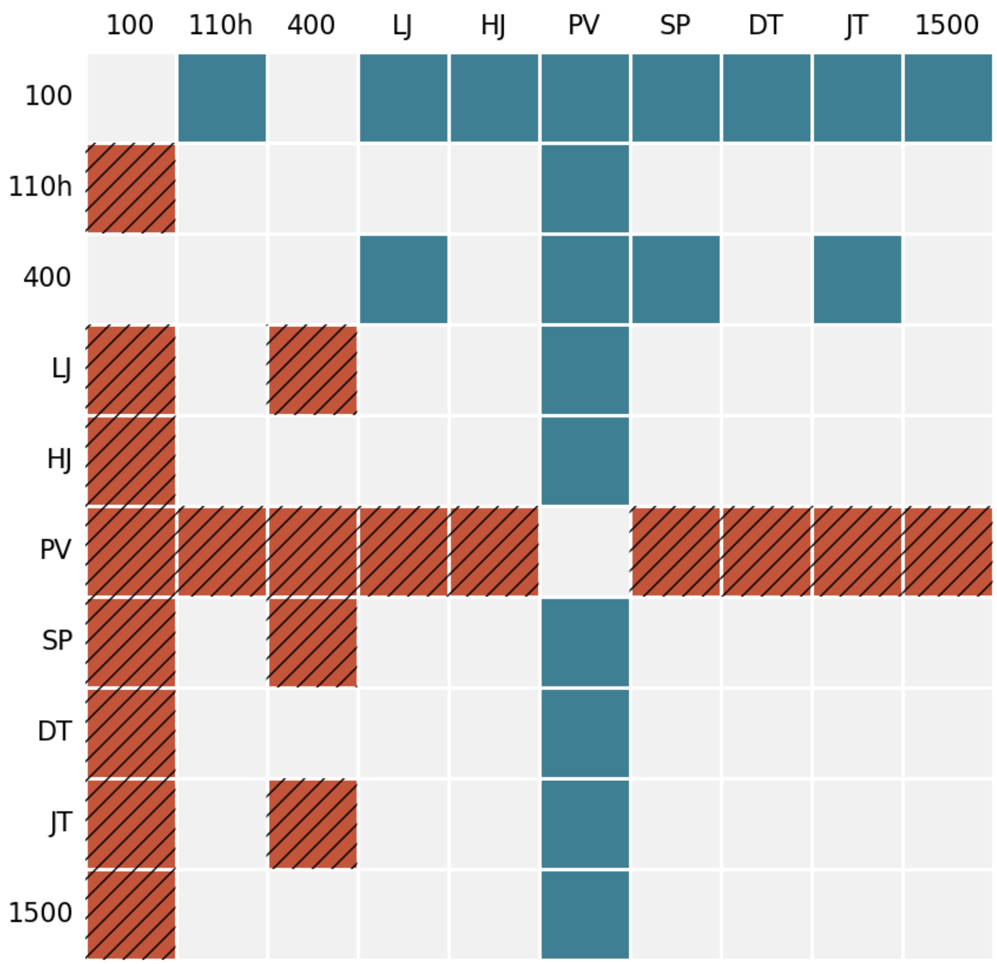


Figure 3: Matrix diagram showing significant differences between FSB coefficients, inverse link . A hatched orange square indicates that the entirety of the bootstrapped confidence interval of the difference between the row coefficient and column coefficient is below zero, indicating the column coefficient is greater. A teal square indicates that the entire confidence interval is above 0, meaning that the row coefficient is greater. A grey square indicates that this interval contains 0, which means that neither one is significantly greater than the other.

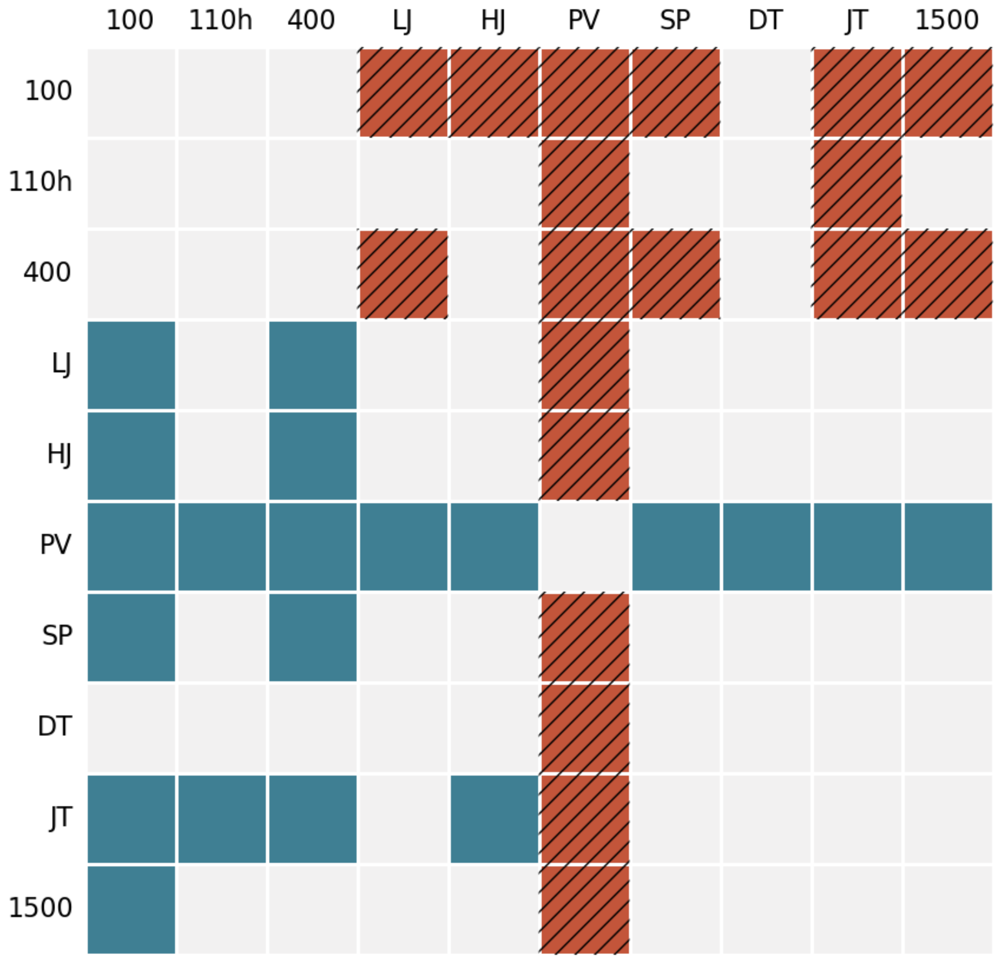
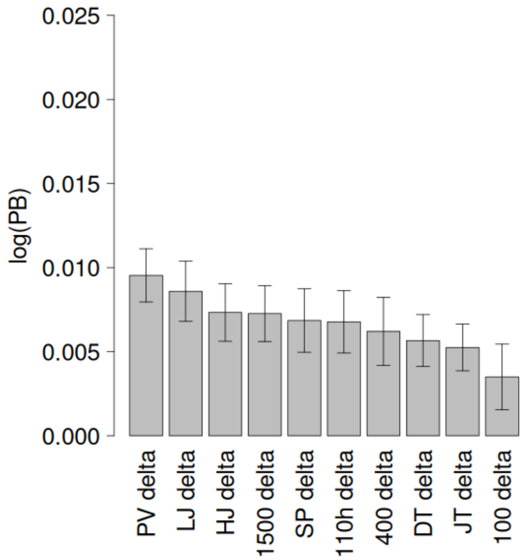
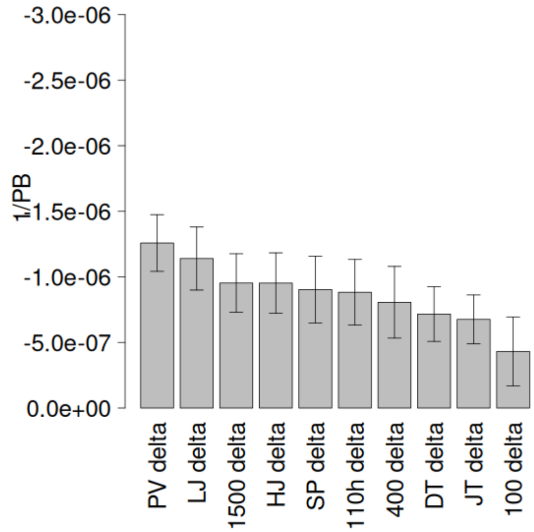


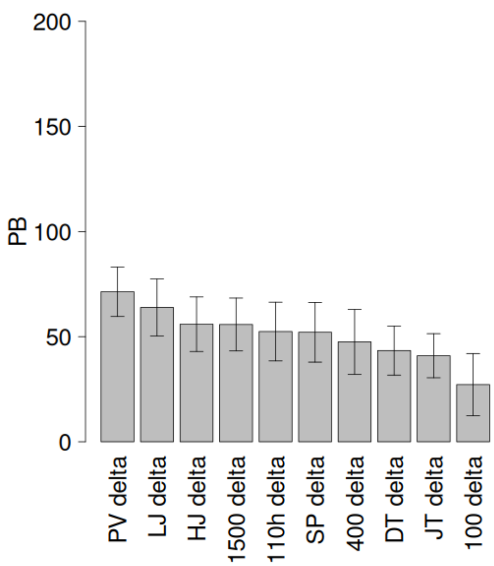
Figure 4: Matrix diagram showing significant differences between FSB coefficients, identity link. A hatched orange square indicates that the entirety of the bootstrapped confidence interval of the difference between the row coefficient and column coefficient is below zero, indicating the column coefficient is greater. A teal square indicates that the entire confidence interval is above 0, meaning that the row coefficient is greater. A grey square indicates that this interval contains 0, which means that neither one is significantly greater than the other.



(a) Log Link



(b) Inverse Link



(c) Identity Link

Figure 5: Absolute magnitudes of delta coefficients, sorted by absolute magnitude from highest to lowest. Bars indicate confidence interval.

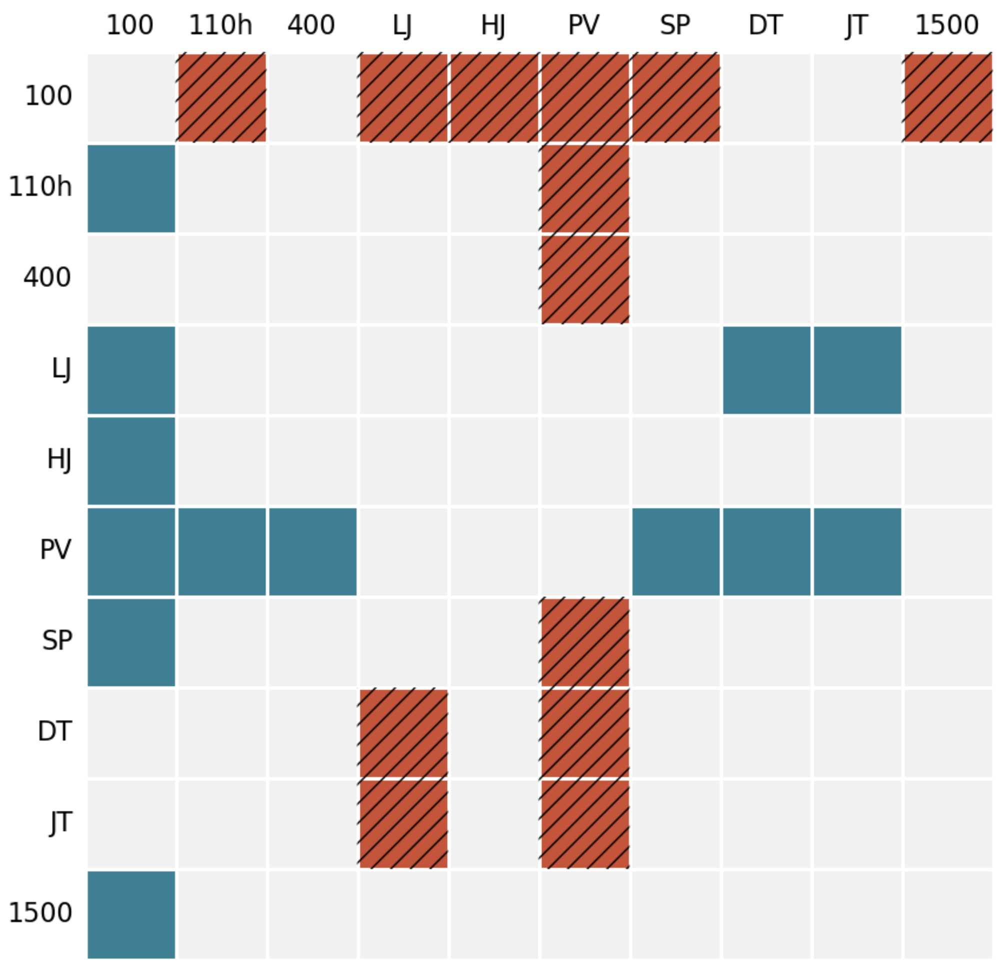


Figure 6: Matrix diagram showing significant differences between delta coefficients, log link. A hatched orange square indicates that the entirety of the bootstrapped confidence interval of the difference between the row coefficient and column coefficient is below zero, indicating the column coefficient is greater. A teal square indicates that the entire confidence interval is above 0, meaning that the row coefficient is greater. A grey square indicates that this interval contains 0, which means that neither one is significantly greater than the other.

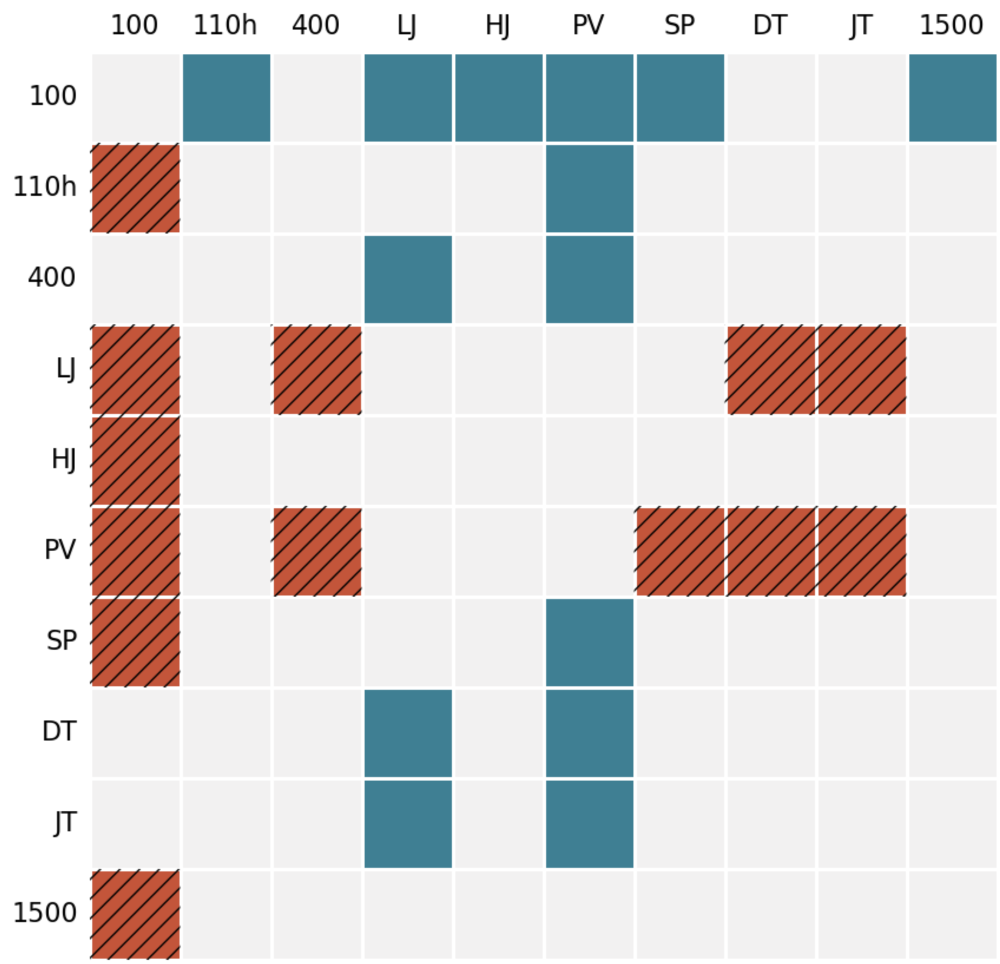


Figure 7: Matrix diagram showing significant differences between delta coefficients, inverse link. A hatched orange square indicates that the entirety of the bootstrapped confidence interval of the difference between the row coefficient and column coefficient is below zero, indicating the column coefficient is greater. A teal square indicates that the entire confidence interval is above 0, meaning that the row coefficient is greater. A grey square indicates that this interval contains 0, which means that neither one is significantly greater than the other.

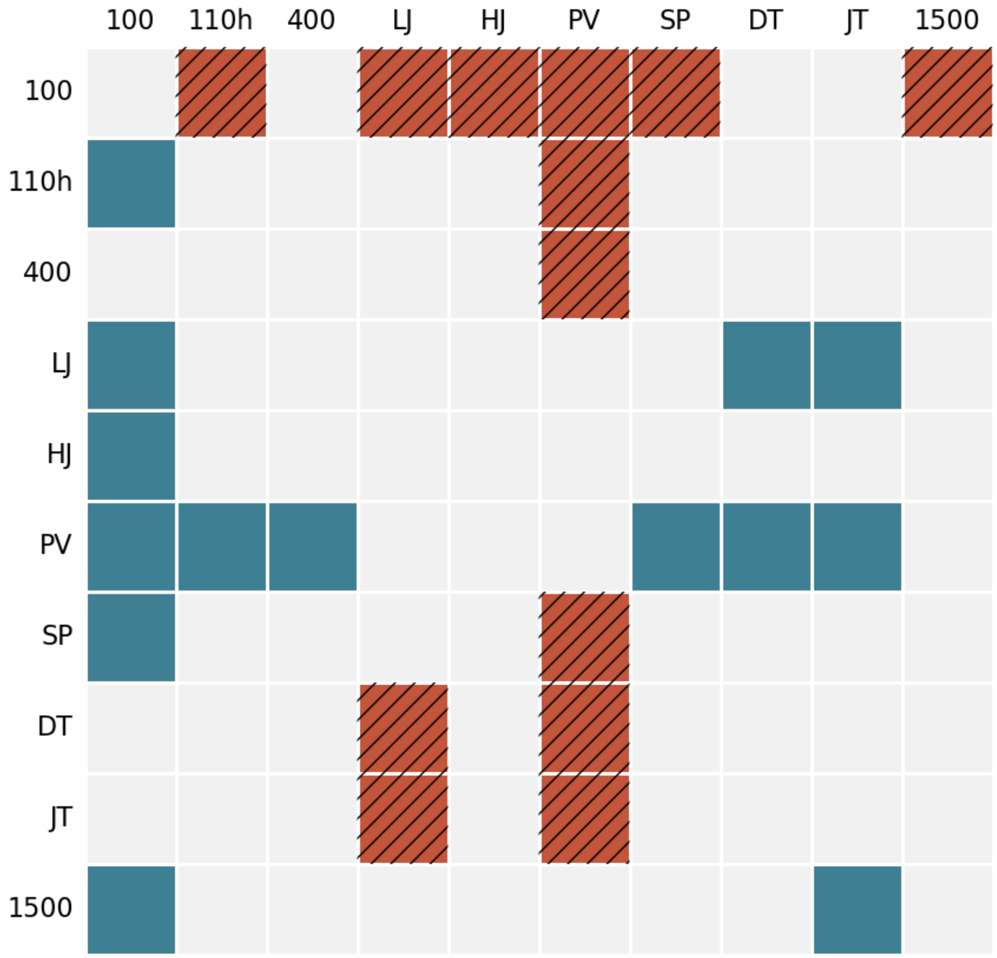


Figure 8: Matrix diagram showing significant differences between delta coefficients, identity link. A hatched orange square indicates that the entirety of the bootstrapped confidence interval of the difference between the row coefficient and column coefficient is below zero, indicating the column coefficient is greater. A teal square indicates that the entire confidence interval is above 0, meaning that the row coefficient is greater. A grey square indicates that this interval contains 0, which means that neither one is significantly greater than the other.

Supporting Information

The following is a summary of the raw, unmodified data provided in `SupData.csv`. This data was subsequently filtered to select eligible athletes, whose marks were then transformed, scaled, and analyzed as described in "Methods"".

"Competitor" through "Men 110h":

Competitor	Men.100	Men.110h
Length:12762	Min. :10.12	Min. :13.36
Class :character	1st Qu.:11.05	1st Qu.:14.79
Mode :character	Median :11.25	Median :15.20
NA	Mean :11.26	Mean :15.26
NA	3rd Qu.:11.46	3rd Qu.:15.66
NA	Max. :14.37	Max. :23.55

"Men 400" through "Men HJ":

Men.400	Men.LJ	Men.HJ
Min. :45.00	Min. :4.140	Min. :1.510
1st Qu.:49.67	1st Qu.:6.660	1st Qu.:1.850
Median :50.66	Median :6.900	Median :1.920
Mean :50.76	Mean :6.899	Mean :1.915
3rd Qu.:51.73	3rd Qu.:7.140	3rd Qu.:1.980
Max. :59.67	Max. :8.450	Max. :2.280

"Men PV through Men's DT"

Men.PV	Men.SP	Men.DT
Min. :2.550	Min. : 7.26	Min. : 2.00
1st Qu.:4.110	1st Qu.:11.81	1st Qu.:35.02
Median :4.400	Median :12.80	Median :38.43
Mean :4.387	Mean :12.82	Mean :38.44
3rd Qu.:4.650	3rd Qu.:13.82	3rd Qu.:41.84
Max. :5.600	Max. :17.78	Max. :55.87

"Men JT" through "Overall Score":

Men.JT	Men.1500	Overall.Score
Min. :12.10	Length:12762	Min. :6400
1st Qu.:47.23	Class :character	1st Qu.:6766
Median :51.86	Mode :character	Median :7152
Mean :52.10	NA	Mean :7207
3rd Qu.:56.89	NA	3rd Qu.:7582
Max. :79.05	NA	Max. :9126

Note: raw marks for the men's 1500m race are provided by world athletics in the format "4:36.7", for example. Therefore, they are represented natively as text rather than numerical values. This file is provided as `SupData.csv` via the submission portal.

Cover Figures

...