Supplementary II: Methods

FOOTBALL DATA SIMULATION

As recommended in O'Kelly, et al. [1], a study protocol was developed before initiation of simulations and analyses. Our methodology was focused on a causal research setting; however, the methods may also be applied in predictive research.[2] Simulation steps 1–4 detailed below are illustrated in online supplemental file 1 figure S1.

Step 1 Preparing data

We constructed different relationships between training load and injury based on a dataset of Norwegian Premier League male football players followed for 323 days (n = 36, mean age 26 years [Standard Deviation 4]). Training load was measured daily with the session Rating of Perceived Exertion (sRPE)[3]: the duration of the activity in minutes multiplied by the player's perceived intensity of the activity on a scale from 0 to 10. The players reported intensity and duration after completion of each training session or match,[4] using a mobile application (Athlete Monitoring, Moncton, Canada). The mean answering time was 0.01 days (SD = 0.2); 99% of prompts were answered within the same day, and the longest answering time was 4 days. Of 4 871 prompts, 650 (13%) Rating of Perceived Exertion observations were missing.[5] The relative training load from one day to the next was calculated with the symmetrized percentage change (%ΔsRPE).[6]

The most common study design in training load and injury risk studies is one team of athletes followed for one season.[7] By rough estimate, a football team suffers on average 40 injuries per team per season, not counting recurrent injuries.[8] The association between training load and injury is likely to be small to moderate,[9] therefore, one team followed for one season is unlikely of sufficient power to detect a relationship accurately,[10] and in most cases, studies will focus on a particular injury type, i.e. hamstring injury. We therefore simulated a medium-to-large-sized study: 250 participants (10 football teams), followed for a season (300 days).

Step 2 Simulating time-to-event data

We simulated injuries under different relationship scenarios with the sampled training load. For simplicity, only one injury was simulated per individual. This scenario may be unrealistic, as sports injuries may be sustained multiple times.[11] The methods for modelling training load considered in this study can, however, also be used with more complex statistical models for recurrent events.[12] The risk of injury at any given time was predetermined with a time-to-event Cox regression model with one covariate:

$$h(t) = h_0(t) * \exp(\beta x) \qquad \textit{Eq. 1}$$

Where h_0 is the baseline hazard, and $h(t)$ is the hazard at timepoint t. The timepoint at which an individual could be censored was drawn at random from a uniform distribution ranging from 0 to 600. Here, x represents the absolute training load, but it can be replaced with the relative training load, $\%\Delta x$. The coefficient β was the result of a bidimensional function on both the magnitude of the training load x, and the distance in time, the time lag l, from the timepoint t. We can write this more accurately:

$$h(t) = h_0(t) * \exp(s(x_t, \dots, x_{t-l}, \dots, x_{t-L})) \qquad \textit{Eq. 2}$$

Here, the function s describes the relationship between training load x and the hazard of injury, measured over the lag interval $l = 0, \dots, L$ where L is the maximum lag. We denoted $l = 0$ to be the current day (Day 0), and the max lag was set at $L = 27$. This corresponds to 28 days (4 weeks).

The s function, $s(x_t, \dots, x_{t-L})$, can be defined in multiple ways.[13] We simulated s to be the cumulative sum of both a function on the magnitude of training load, the variable function $f(x)$, and a function on the distance in time from the current day, the lag function $w(l)$. This can be represented by:

$$s(x_t, \dots, x_{t-L}) = \sum_{l=0}^{L} f(x) \cdot w(l) \qquad \textit{Eq. 3}$$

The shape of the relationship between the absolute training load and injury risk was simulated to be J-shaped (online supplemental file 1 figure S2A).[14] Under this assumption, the lowest point of risk was intermediate levels of training load. The highest was under high levels of training load. The variable function $f(x)$ was:

$$f(x) = \begin{cases} ((600 - x)/200)\text{^}1.5/10, & x < 600 \\ ((x - 600)/200)\text{^}3/30), & x \geq 600 \end{cases}$$

Where x was measured with the sRPE. For the relative training load, we simulated a linear relationship with injury risk (figure S2C). Higher loads on the current day compared to load on the previous day increases risk, whilst lower loads on the current day compared with the previous day reduces risk[15]:

$$f(\%\Delta x) = 0.009 * \%\Delta x$$

Here, $\%\Delta x$ was the symmetrized percent change from the previous day, ranging from -100% to 100%.

To compare methods ability to discover different time-dependent effects, the lag function $w(l)$ was defined in four different scenarios.

Constant. Across 4 weeks, the effect of training load has a constant effect each day (online supplemental file 1 figure S3A). Thereafter, training load has no effect. This was an overly simplistic base scenario.

$$w(l) = 0.8$$

Decay. Across 4 weeks, the effect of training load gradually decays for each day (figure S3B).[16] Thereafter, training load has no effect. This was hypothesized as a likely scenario if past training load has a direct effect on injury risk.

$$w(l) = \exp\left(-\frac{l}{100}\right)$$

Exponential decay. On the current day, training load has the highest risk of injury. The effect of training load drops exponentially the past 4 weeks (figure S3C). Thereafter, training load has no effect. This was hypothesized as a likely scenario if past training load has an indirect effect on injury risk.

$$w(l) = \exp\left(-\frac{l}{10}\right)^2$$

Direct, then inverse. Training load values on the current week (acute) increases risk of injury, whilst the training load values three weeks before the current week (chronic) decreases risk of injury (figure S3D)[17] Thereafter, training load has no effect. This hypothesis has recently been challenged.[18 19] Nevertheless, to ensure that modelling methods can uncover this relationship should it be true, we opted to include it regardless. The theory depends on chronic load amount as a surrogate measure for fitness, and acute load amount a surrogate measure for fatigue.[15] High loads relative to the previous time period are thought to increase risk, while low loads relative to the previous time period decrease risk: a linear relationship.[15 20 21] Therefore, for this time-lag scenario, we simulated a linear relationship with the absolute training load, and the relative load was not considered,

$$w(l) = \begin{cases} \exp\left(-\frac{l}{10}\right)^2, & l \leq 6 \\ -\exp\left(\frac{l}{50}\right)^2, & l > 6 \end{cases}$$

The relationships constant, decay and exponential decay were used both for the absolute training load and for the relative training load. The “Direct, then inverse” relationship was only simulated for the absolute training load exposure. For this time-lag scenario, and for this time-lag scenario only, we simulated a linear relationship with the absolute training load (online supplemental file 1 figure S2B):

$$f(x) = 0.0009 * x$$

In summary, seven different relationships between training load and injury risk were simulated (figure 1–2). In a pilot of 100 simulations for each of the seven scenarios, the mean number of simulated injuries for 25 participants (a football team) was 18.7 per season; reasonably realistic of a small-to-moderate effect between training load and a specific injury type (i.e. a study on hamstring injury).

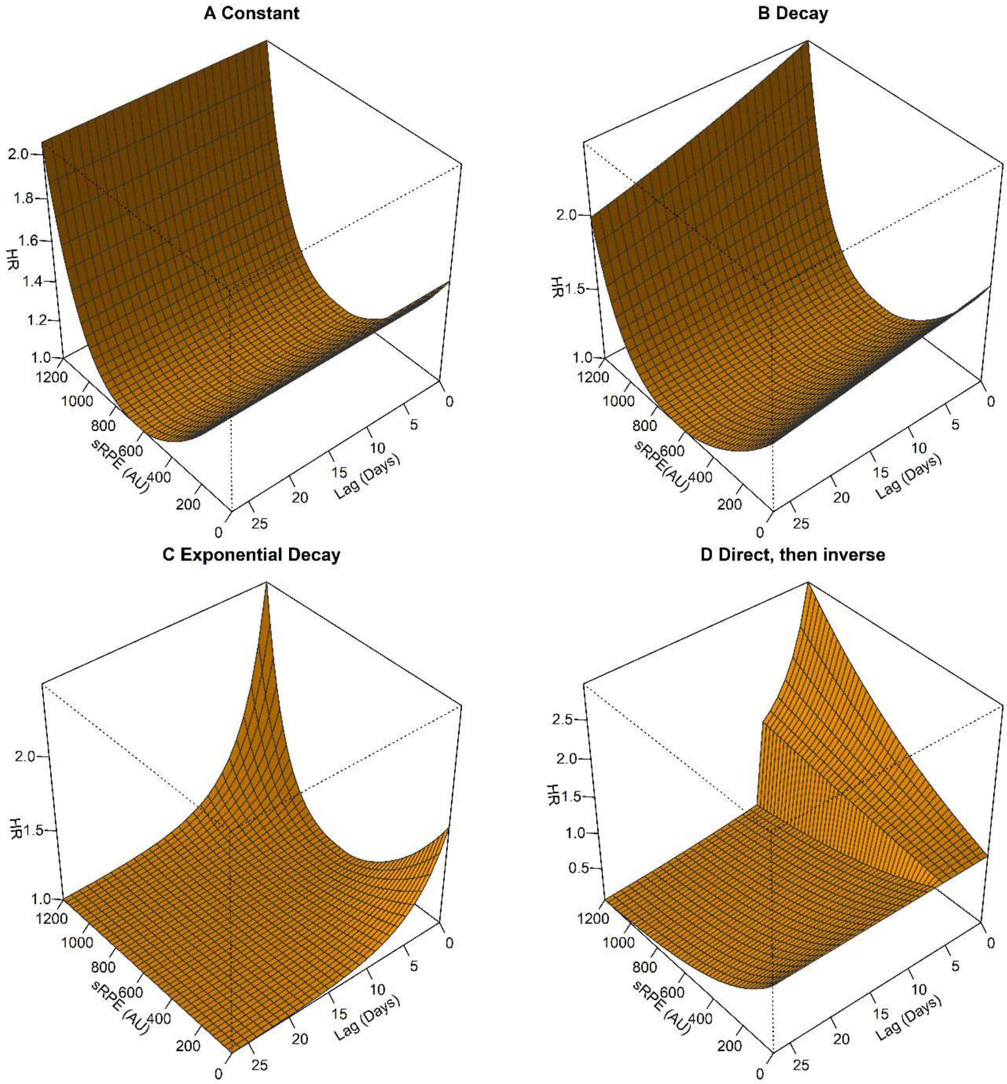

Figure 1. The four simulated relationships between absolute training load and injury risk. The relationships are a combination of the J-shaped function on the absolute training load exposure (online supplemental file 1 figure S2A) and the different functions on the time since training load was sustained (figure S3). Training load is measured with the session Rating of Perceived Exertion (sRPE), shown on the X-axis. The time since the current day (Day 0) is shown on the Y-axis, where 0 is the current day and 27 is the 27th day before the current day. On the Z-axis, the risk of injury is measured with the Hazard Ratio (HR), where HR > 1 indicates an increased risk, and HR < 1 indicates a decreased risk. The four risk shapes are (A) Constant, where the J-shaped risk of training load is constant over time; (B) Decay, where the effect-size of the J-shaped effect of training load is at its highest on the current day (Day 0) and is reduced linearly for each lag day back in time; (C) Exponential Decay, where the J-shaped risk of training load is at its highest on the current day (Day 0) and is reduced exponentially for each lag day back in time; (D) Direct, then inverse; where training load linearly increases injury risk during the current week (Day 0–Day 6), but linearly decreases injury risk thereafter. This was the shape simulated with a linear model on the absolute training load (figure S2B). Training load had no effect after the 27th lag day (4 weeks) in all four scenarios (not shown).

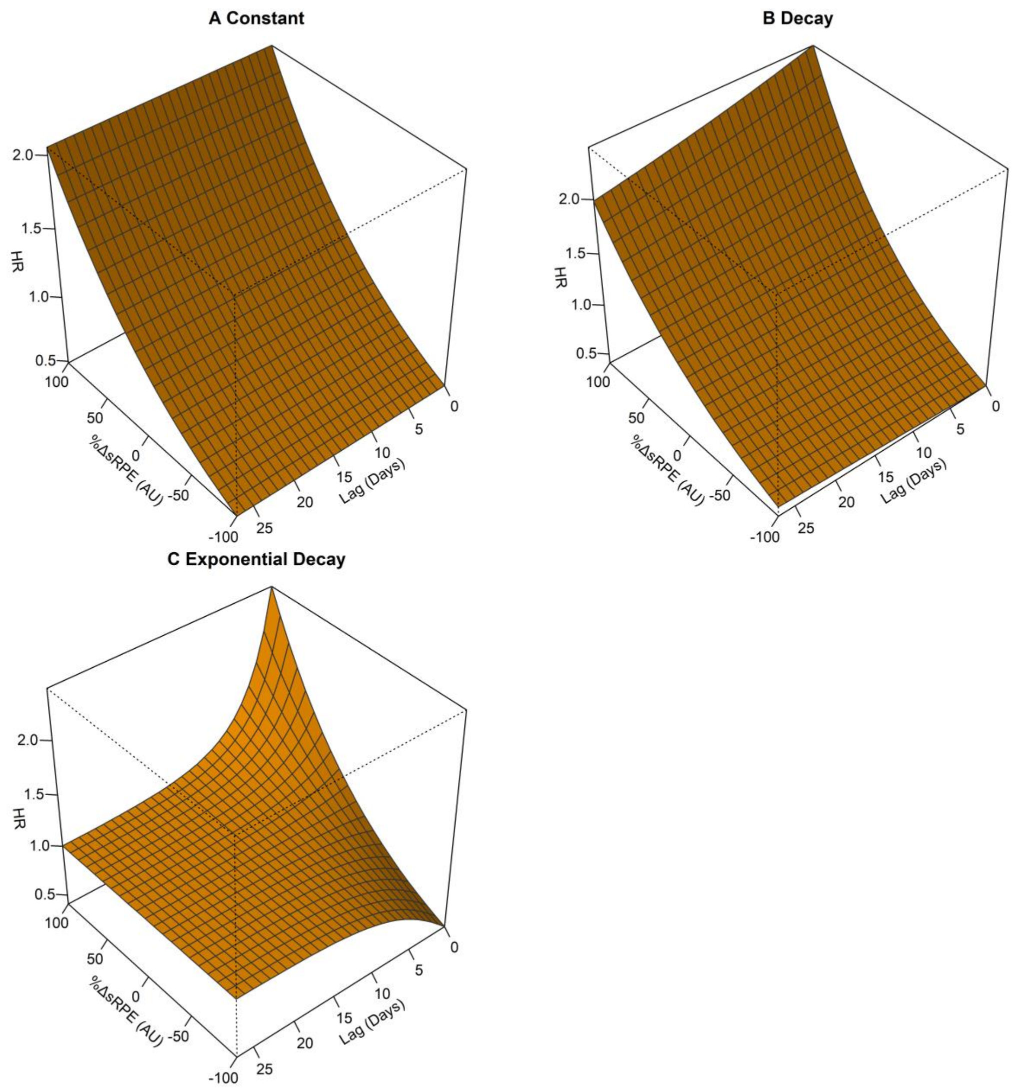

Figure 2. The three simulated relationships between relative training load and injury risk. The relationships are a combination of the linear function on the relative training load exposure (online supplemental file 1 figure S2C) and the different functions on the time since training load was sustained (figure S3). Relative training load is measured with the symmetrized percentage change (%Δ) in session Rating of Perceived Exertion (sRPE), shown on the X-axis. The time since the current day (Day 0), the number of lag days is shown on the Y-axis, where 0 is the current day and 27 is the 27th day before the current day. On the Z-axis, the risk of injury is measured with the Hazard Ratio (HR), where HR > 1 indicates an increased risk, and HR < 1 indicates a decreased risk. The four risk shapes are (A) Constant, where the linear risk of relative training load is constant over time; (B) Decay, where the effect size of the linear effect of relative training load is at its highest on the current day (Day 0) and is reduced linearly for each lag day back in time; (C) Exponential Decay, where the linear risk of training load is at its highest on the current day (Day 0) and is reduced exponentially for each lag day back in time. Training load had no effect after the 27th lag day (4 weeks) in all three scenarios (not shown).

Step 3 Modelling the time-dependent effect of training load on injury risk

Different methods of modelling training load were compared in their ability to uncover the seven predetermined relationships between training load and injury risk. A Cox regression model (Eq. 1) was used to estimate the relative risk of injury, where training load, x or $\%\Delta x$, was modified or modelled in three different ways for the absolute training load, and three different ways for the relative training load.

We chose the most frequently used methods in training load and injury research,[22-24] methods proposed as potential alternatives,[16 25] and a method developed to handle similar challenges in epidemiology.[26 27]

In the Cox regression model, regardless of method used to modify the absolute training load, the training load was modelled with a quadratic term under all time-lag scenarios except for the "Direct, then inverse", where a linear term was used. This was done to ensure methods were compared under the same conditions. Here, we assumed that a given researcher would have performed a sensitivity analysis before-hand to determine the need for a linear vs. non-linear shape.

A linear relationship was assumed between relative training load and injury risk, regardless of method used to modify the training load.

Absolute training load

Rolling average

Despite past critiques,[28] the rolling average (RA)[29] was the most frequently used method to account for the cumulative effects of training load in recent reviews.[23 30] Training load and injury risk studies that employ more advocated methods[16] still calculate the RA alongside the other calculations.[31-34] We therefore included this method in our comparison. For training load denoted x, the moving average RA is defined by:

$$RA = \frac{x_{k-n+1} + x_{k-n+2} + \cdots + x_k}{n}$$

Where n is the size of the time-lag window, in this study, 28 days. k denotes the last value in the time-lag window for an individual. For the first window, k = 28, for the second window, $k = 29$, and so on, up until the final window, k = 300. For each window, the first value is removed from the calculation, and the next value is added. For example, the first rolling average calculation is:

$$RA_1 = \frac{x_1 + x_2 + \cdots + x_{28}}{28}$$

The second rolling average calculation is:

$$RA_2 = \frac{x_2 + x_3 + \cdots + x_{29}}{28}$$

This sliding window of calculation can thus be generalized to:

$$RA_{today} = RA_{yesterday} + \frac{1}{n}(x_{k+1} - x_{k-L+1})$$

The method is intuitive and simple to calculate. An advantage is that it can be calculated on incomplete time-windows, given that n is defined as the number of training load values in the time sequence so far. For comparability with other methods, however, we calculated RA only from the 28th value and so on. The disadvantage is that rolling averages assume that training loads further back in time, and more recent training loads, contribute equally to injury risk.[16] The method provides no flexibility in the size or direction of effect for different time-lags.[35]

Exponentially weighted moving average

The exponentially weighted moving average (EWMA) is an extension of the rolling average. It accounts for the assumption that training load values further back in time contribute less to injury risk than training loads closer in time to the current day.[16] It has been recommended as an improvement over the rolling average, [16 36] and has been used in training load and injury risk studies since.[24 30 33] For training load denoted x, EWMA is:

$$EWMA_{today} = x_{today} + \lambda + ((1 - \lambda) + EWMA_{yesterday})$$

Where λ represents the decrease in effect depending on distance in time, by number of days n, up to a maximum of $n = 28$:

$$\lambda = \frac{2}{n + 1}$$

This choice of lambda is the same as in Williams, et al. [16] and Moussa, et al. [25].

A disadvantage of the EWMA is that a full window (28 days) must be completed before the calculation of the first EWMA. Any injuries sustained in this period are therefore not included in the analysis of injury risk. In addition, EWMA is constrained to an exponential weight only, and it cannot be calculated in the presence of missing values.[25]

Robust exponential decreasing index

The Robust Exponential Decreasing Index (REDI) has recently been proposed as an alternative over the EWMA,[25] and had improved performance in a training load and injury risk study.[37] For the lag interval $l = 0, \dots, L$ where $l = 0$ is the current day, and L is the maximum lag 27, we can determine a vector of coefficients for each lag. Then, multiply the coefficients with the training load at each lag and sum these weighted training load values.

$$\text{Weighted x} = \sum_{l=0}^{\text{L}} \alpha_l^{\lambda} * \text{x}_\text{l}$$

The coefficient, α_l^{λ} is determined as follows:

$$\alpha_l^\lambda = \begin{cases} 0 & if\ x\ is\ missing \\ \exp(-\lambda * l) & if\ x\ is\ not\ missing \end{cases}$$

The λ weight has to be specified by the user, same as the EWMA method. The weighted training load values are then divided by the sum of the weights:

$$REDI = \frac{\text{Weighted x}}{\sum_{l=0}^{\mathrm{L}} \alpha_l^\lambda}$$

The lower the lambda ($\lambda \rightarrow 0$), the greater the impact from past training load values. We chose lambda = 0.1 as it was the highest lambda value where training load on the 27th lag day still contributed to the cumulative effect.[25] Coincidentally, it was the same as used in Moussa, et al. [25], and is closest in behavior to the EWMA.

REDI is robust to missing data in training load, and like the rolling average, it can be calculated on incomplete time-windows. In addition, it may be more flexible than the EWMA in that the choice of lambda can fine-tune the weights to a specific sport or setting.[25]

Distributed lag non-linear model

In environmental epidemiology, modelling long-term effects – such as pollution or radon-exposure – is a common challenge. Although not entirely applicable to the challenges with training load, they do share the complexities of being long-term, weak-to-moderate protracted time-varying effects.

To recap, the relative risk of injury is considered to be the combined result of 1) the magnitude of exposure to training load, known as the exposure-response relationship, and 2) the distance in time from the current day (Day 0), the lag-response relationship.

To handle such effects, Bhaskaran, et al. [26] suggested using a so-called distributed lag model, a method initially developed in econometrics[38] and later applied to epidemiology.[39]

With Eq. 2, we explained how the β-coefficient for training load can be a result of the s function, $s(x_t, \ldots, x_{t-L})$. In a distributed lag model, the effects from the lag-response relationship is modelled with the lag-response function $w(l)$:

$$s(x_t, \ldots, x_{t-L}) = \sum_{l=0}^{L} x_{t-l} w(l)$$

When $w(l)$ is a constant function, this is equivalent to the rolling average.[13] Distributed lag models has been implemented in environmental epidemiology to handle cumulative, time-dependent effects.[26 40] The downside is the data-driven exploration of cut-offs,[35] and the assumption of a linear relationship between exposure, lag and response.[26]

To account for these issues, Bhaskaran, et al. [26] recommended using polynomial or splines to explore the long-term pattern in so-called Distributed Lag Non-linear Models (DLNM).

This has been applied to time-to-event data in medicine.[41 42] DLNMs allow non-linear modelling of the combined effect of the exposure-response and the lag-response relationships: the exposure-lag-response relationship.[27] The function s can be defined by crossing the variable function $f(x)$ and the lag function $w(x,l)$ and thus produce a bi-dimensional exposure-lag-response function $f(x) \cdot w(x,l)$:

$$s(x_t, \ldots, x_{t-L}) = \sum_{l=0}^{L} f(x) \cdot w(x_{t-l}, l)$$

The exposure-response function $f(x)$, the function on the absolute training load, must be specified by the user. In the Cox regression model, $f(x)$ was modelled with a quadratic term, except for the "Direct, then inverse" time-lag scenario, where a linear term was used instead; same as for the other methods. The lag-response function $w(x,l)$ is the function for the time-dependent effect, and must also be specified by the user. Here, it was modelled with restricted cubic splines using 3 knots under all scenarios, since splines can explore non-linear shapes.[14] For a gentle introduction to DLNMs, see Gasparrini [13]. For more extensive mathematical exploration, see Gasparrini [27].

DLNM is a method which models, rather than modifies, training load. Therefore, no discarding of data, choice of time-blocks, or aggregation of training load values is necessary, and so, all information in the raw data is retained. Another advantage is that DLNM is flexible in the modelling of the exposure-response and the lag-response functions, both of which may be modelled with polynomials or splines at the user's discretion. This allows the exploration of non-linear and complex time-lag effects. On the other hand, modelling complex time-lag effects may require larger sample sizes, and model specification requires subjective choice.[13]

Relative training load

Week-to-week percentage change

In training load studies, it is common to divide the data into blocks of time.[43 44] The weekly sRPE is calculated by summing the daily sRPEs.[34] The percentage difference can then be calculated on the difference in sRPE between the current week and the previous week.[45 46] We included this method in the comparison as the most basic method of calculating relative training load. The percentage difference has a few disadvantages,[6] one being that it cannot be calculated when the denominator is zero. We therefore opted for the symmetrized percentage change, which has improved mathematical properties.[6] This calculation can be represented by:

$$\%\Delta \mathrm{W} = \frac{W_k - W_{k-1}}{W_k + W_{k-1}} * 100$$

Where k is the current week. In the same manner as the moving average, the week-to-week percentage change calculation moves iteratively from one week to the next.

The week-to-week percentage change is simple to calculate. Any injuries suffered in the first six days must be discarded before calculation of the first percentage difference. However, this is a small amount of data compared to some of the other methods compared. The main disadvantage is that it does not consider training load values further back in time than the previous week, and the time-block of a week may be unreasonable for many sports.[47]

Acute: Chronic Workload Ratio
In 2016, Blanch and Gabbett [17] introduced the Acute: Chronic Workload Ratio (ACWR), which is the most frequently used method of modifying training load before analysing the effect of training load on injury risk.[22 48] The training load on the current week (Day 6 up to Day 0) is considered the "acute" training load. The "chronic" training load is typically defined as the rolling average of the current week and the previous three weeks (Day 27 up to Day 0), known as the or 7:28 ACWR. As shown in,[49] the basic ACWR calculation is:

$$\text{ACWR} = \frac{Acute\ Week}{Chronic\ Weeks * 0.25} = \frac{W_k}{(W_{k-3} + W_{k-2} + W_{k-1} + W_k) * 0.25}$$

Where k is the current week. In the same manner as the rolling average, the traditional ACWR calculation moves iteratively from one week to the next. We calculated ACWR from one day to the next, a calculation less wasteful of data.[47]

ACWR can be calculated in many different ways.[22 23] The time windows for the acute and chronic periods are at the user's discretion.[22 47] The acute load is typically the sum of training load exposures on the current week, but the chronic load can by calculated by either the rolling average or the EWMA.[23 36 50] Finally, in the traditional ACWR, the acute load is included in the denominator. This is known as the "coupled" ACWR. The "uncoupled" ACWR – where the acute load is *not* included in the denominator – has been recommended as a more concrete measure of the change in training load.[18 21] For this simulation study, we chose the coupled 1-week absolute sum: 4 week rolling average ACWR, the most common form of calculation.[23]

The advantage of the ACWR is addressing the potential effect of the relative training load, while also accounting for past exposure. The properties of the ACWR has been explored extensively, with multiple critiques.[18 19 22 23 51] Like EWMA, ACWR needs a completed time window before the first calculation.

Distributed lag non-linear model
The ability of the distributed lag non-linear model (DLNM) to uncover the effect of relative training load was also assessed. The exposure-response function $f(\%\Delta x)$ was assumed to be linear, the same assumption as for the ACWR and week-to-week percentage change. The lag-response function $w(x, l)$ was modelled with restricted cubic splines using 3 knots under all scenarios.

Step 4 Calculating performance measures
Metrics for comparing the model fit, accuracy and certainty of the models were calculated in the final step.

Root-Mean-Squared Error
For a measure of accuracy, we calculated the difference between the predicted cumulative

hazard $\hat{\theta}$ and the true cumulative hazard θ used to simulate the survival data for a range of training load values, the absolute bias. The main performance measure was the Root-Mean-Squared Error (RMSE), calculated by:

$$RMSE = \sqrt{mean\left(\left(\hat{\theta} - \theta\right)^2\right)} = \sqrt{mean(bias^2)}$$

RMSE is a combined measure of accuracy and precision, where the lower the RMSE, the better the method.[13] The scale of the RMSE depends on the scale of the coefficients in question, and it is therefore only interpretable by comparing values in the same analysis – the values cannot be interpreted in isolation.[52]

For the relative training load, the ACWR and the week-to-week percentage change methods modified the training load values to a different scale than the one used to simulate the data. The RMSE for the predicted vs. true cumulative hazard, a measure of external validation, could therefore not be calculated for each level of percentage change in training load. Therefore, we also calculated RMSE on the predicted injury value vs. the observed value (the model residuals), as an internal validation:

$$RMSE_{Internal} = \sqrt{mean(residuals^2)}$$

Model fit
Model fit was measured by Akaike's Information Criterion (AIC) which has shown to be more appropriate than BIC for comparison of time-lag models.[27] The AIC can be used to compare non-nested models,[53-55] but the AIC is not comparable if models are run on different sample sizes.[53] Since some methods – EWMA, ACWR – required the completion of a full time period before first calculation, the first 27 rows were removed from the dataset for all methods before fitting the Cox regression model to ensure comparability of the AIC.

Coverage
Coverage was calculated as the proportion of 95% confidence intervals that contained the true value. Average width (AW) of the 95% confidence intervals was also calculated, as a measure of statistical efficiency.

Number of simulations

Using formulas listen in Morris, et al. [52], accepting a Monte Carlo Standard Error of no more than 0.5, the number of simulations needed for an accurate determination of coverage was:

$$n_{Coverage} = \frac{E(Coverage)(1 - E(Coverage)}{(Monte\ Carlo\ SE_{req})^2} = \frac{95 * 5}{0.5^2} = 1\ 900$$

The number of simulations needed for an accurate estimate of bias was calculated by:

$$n_{sim} = \frac{s^2}{0.5^2}$$

Where s is the sample variance of bias.[52] For an estimation of variance, a pilot of 200 simulations were run for each constructed relationship. The highest variance in bias was

6.63, and the number of simulations needed to achieve the target MCSE was 176. Since coverage required more simulations to achieve target MCSE, simulation steps 1–4 outlined above were repeated 1 900 times. The mean of each performance measure was calculated across the 1 900 simulations.

IMPLEMENTATION IN A HANDBALL COHORT

The distributed lag non-linear model (DLNM) was implemented on an observed handball cohort to illustrate how it can be used in practice. To explore the potential for a time-dependent, cumulative effect of training load on injury risk, we chose the Norwegian elite youth handball data. The data was a cohort of 205 elite youth handball players from five different sport high schools in Norway (36% male, mean age: 17 years [SD: 1]) followed through a season from September 2018 to April 2019 for 237 days.[56]

RPE and duration was collected from the players after each training and match, from which daily sRPE was determined.[56] Timeliness was relatively poor; 53% of activity prompts were answered on the same day, and the mean number of days from prompt to reply was 0.7 (SD = 1.6). Of 47 651 activity prompts, 64% were missing, likely under the missing at random or missing not at random mechanism.[57] Missing sRPE data had previously been imputed with multiple imputation using predicted mean matching,[5] before the data were anonymized.[14] All non-derived variables were used to predict imputed values, including age, sex, player position, training activity type among others. The response variable, injury, was also used to predict imputed values,[58] but was not itself imputed before analysis.[59] The duration and RPE variables, the factors from which sRPE is derived, were not included in the imputation model for predicting sRPE.[5] The number of imputed datasets, five, is recommended in most cases.[60] The observed distribution was maintained in the imputed values; therefore the imputation was deemed valid.[14] Although the poor data quality rendered the handball data unsuitable for a study of causal inference, it had a sufficient number of injuries for the current methodology study (n = 472), and previously showed a potential non-linear relationship between training load and injury risk.[14]

The handball players reported whether they had “no health problem”, “a new health problem”, or “an exacerbation of an existing health problem” each day. Any response of “a new health problem” was considered an injury event in the current study. Players were encouraged to report all physical complaints, irrespective of their consequences on sports participation or the need to seek medical attention.[61]

A Cox regression model was run with injury (yes/no) as the outcome and the DLNM of sRPE as the exposure of interest.[62] DLNM combines a dose-function on the magnitude of sRPE, and a lag-function on the distance since Day 0, up to lag 27 (4 weeks). The dose-function was modelled with a restricted cubic splines with 3 knots.[14] Based on AIC, a linear model was chosen for the lag-function. The Cox model was adjusted for sex and age as potential confounders. A frailty term with a gamma distribution was used to account for recurrent events.[12] The model predictions were visualized to assess the ability of DLNM to explore effects. Predictions from each of the imputed datasets were averaged, then visualized.[63]

DATA TOOLS

The simulations were run on an Intel(R) Core(TM) i7-6700K 4.00GHz CPU, 16 GB RAM computer. All statistical analyses and simulations were performed using R 4.1.2[64] with RStudio version 1.4.1717. A GitHub repository is available with R code and data used in the simulations.[65] PermAlgo was used to simulate survival data.[42 66] The slider package was used for calculations on sliding windows,[67] using zoo[68] for rolling averages and TTR[69] for EWMA. Handling time-lag data and performing distributed lag non-linear models was done with DLNM.[70]

ETHICS

Data collection for both studies were approved by the Ethical Review Board of the Norwegian School of Sport Sciences. They were also approved by the Norwegian Centre for Research Data: Norwegian Premier League football (722773); Norwegian elite youth handball (407930). All participants provided informed written consent. They were all above the age of 15 and parental consent was not required. Ethical principles were followed in accordance with the Declaration of Helsinki,[71] with the exception that the study was not registered in a publicly accessible database before recruitment of the first subject (a violation of principle number 35). Data were anonymised according to guidelines outlined by The Norwegian Data Protection Authority.[72] The datasets cannot be joined.

REFERENCES

1. O'Kelly M, Anisimov V, Campbell C, et al. Proposed best practice for projects that involve modelling and simulation. *Pharmaceutical statistics* 2017;16(2):107-13.
2. Shmueli G. To explain or to predict? *Statistical science* 2010;25(3):289-310.
3. Foster C, Florhaug JA, Franklin J, et al. A New Approach to Monitoring Exercise Training. *Journal of Strength and Conditioning Research* 2001;15:10Y 115.
4. Rønneberg KL. Seasonal training load quantification in men's Norwegian premier league football: Differences in measured external-and internal training load within microcycles and throughout the competition phase. Norwegian School of Sport Sciences, 2020.
5. Bache-Mathiesen LK, Andersen TE, Clarsen B, et al. Handling and reporting missing data in training load and injury risk research. *Science and Medicine in Football* 2021:1-13. doi: 10.1080/24733938.2021.1998587
6. Curran-Everett D, Williams CL. Explorations in statistics: the analysis of change. *Advances in physiology education* 2015;39(2):49-54.
7. Windt J, Ardern CL, Gabbett TJ, et al. Getting the most out of intensive longitudinal data: a methodological review of workload–injury studies. *BMJ open* 2018;8(10):e022626.
8. Ekstrand J, Krutsch W, Spreco A, et al. Time before return to play for the most common injuries in professional football: a 16-year follow-up of the UEFA elite Club injury study. *British journal of sports medicine* 2020;54(7):421-26.
9. Bahr R. Why screening tests to predict injury do not work—and probably never will…: a critical review. *Br J Sports Med* 2016;50(13):776-80.
10. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II-binary and time-to-event outcomes. *Statistics in medicine* 2019;38(7):1276-96.
11. Meeuwisse WH, Tyreman H, Hagel B, et al. A dynamic model of etiology in sport injury: the recursive nature of risk and causation. *Clinical Journal of Sport Medicine* 2007;17(3):215-19.
12. Ullah S, Gabbett TJ, Finch CF. Statistical modelling for recurrent events: an application to sports injuries. *Br J Sports Med* 2014;48(17):1287-93.

13. Gasparrini A. Modelling lagged associations in environmental time series data: a simulation study. *Epidemiology (Cambridge, Mass)* 2016;27(6):835.
14. Bache-Mathiesen LK, Andersen TE, Dalen-Lorentsen T, et al. Not straightforward: modelling non-linearity in training load and injury research. *BMJ Open Sport & Exercise Medicine* 2021;7(3):e001119. doi: 10.1136/bmjsem-2021-001119
15. Gabbett TJ. The training—injury prevention paradox: should athletes be training smarter and harder? *British journal of sports medicine* 2016;50(5):273-80.
16. Williams S, West S, Cross MJ, et al. Better way to determine the acute: chronic workload ratio? *British journal of sports medicine* 2017;51(3):209-10.
17. Blanch P, Gabbett TJ. Has the athlete trained enough to return to play safely? The acute: chronic workload ratio permits clinicians to quantify a player's risk of subsequent injury. *British journal of sports medicine* 2016;50(8):471-75.
18. Wang C, Vargas JT, Stokes T, et al. Analyzing Activity and Injury: Lessons Learned from the Acute: Chronic Workload Ratio. *Sports Medicine* 2020:1-12.
19. Impellizzeri FM, Tenan MS, Kempton T, et al. Acute: Chronic Workload Ratio: Conceptual Issues and Fundamental Pitfalls. *International Journal of Sports Physiology and Performance* 2020;15(6):907-13.
20. Wang C, Stokes T, Steele R, et al. 118 Application of the acute:chronic workload ratio in children. *British Journal of Sports Medicine* 2021;55(Suppl 1):A47-A48. doi: 10.1136/bjsports-2021-IOC.110
21. Wang C, Stokes T, Vargas JT, et al. Injury risk increases minimally over a large range of the acute-chronic workload ratio in children. *American Journal of Epidemiology* 2021 doi: 10.1093/aje/kwab280
22. Dalen-Lorentsen T, Andersen TE, Bjørneboe J, et al. A cherry tree ripe for picking: The relationship between the acute: chronic workload ratio and health problems. *Journal of Orthopaedic & Sports Physical Therapy* 2021(0):1-26.
23. Wang A, Healy J, Hyett N, et al. A systematic review on methodological variation in acute: chronic workload research in elite male football players. *Science and Medicine in Football* 2021;5(1):18-34.
24. Udby CL, Impellizzeri FM, Lind M, et al. How has workload been defined and how many workload-related exposures to injury are included in published sports injury articles? A scoping review. *Journal of orthopaedic & sports physical therapy* 2020;50(10):538-48.
25. Moussa I, Leroy A, Sauliere G, et al. Robust Exponential Decreasing Index (REDI): adaptive and robust method for computing cumulated workload. *BMJ Open Sport & Exercise Medicine* 2019;5(1)
26. Bhaskaran K, Gasparrini A, Hajat S, et al. Time series regression studies in environmental epidemiology. *International journal of epidemiology* 2013;42(4):1187-95.
27. Gasparrini A. Modeling exposure–lag–response associations with distributed lag non-linear models. *Statistics in medicine* 2014;33(5):881-99.
28. Menaspà P. Are rolling averages a good way to assess training load for injury prevention? *Br J Sports Med* 2017;51(7):618-19.
29. Drew MK, Blanch P, Purdam C, et al. Yes, rolling averages are a good way to assess training load for injury prevention. Is there a better way? Probably, but we have not seen the evidence. *British Journal of Sports Medicine* 2017;51(7):618-19. doi: 10.1136/bjsports-2016-096609
30. Andrade R, Wik EH, Rebelo-Marques A, et al. Is the Acute: Chronic Workload Ratio (ACWR) Associated with Risk of Time-Loss Injury in Professional Team Sports? A Systematic Review of Methodology, Variables and Injury Risk in Practical Situations. *Sports Medicine* 2020:1-23.
31. Nakaoka G, Barboza SD, Verhagen E, et al. The association between the acute: chronic workload ratio and running-related injuries in Dutch runners: a prospective cohort study. *Sports medicine* 2021:1-11.

32. Arazi H, Asadi A, Khalkhali F, et al. Association Between the Acute to Chronic Workload Ratio and Injury Occurrence in Young Male Team Soccer Players: A Preliminary Study. *Frontiers in Physiology* 2020;11
33. Xiao M, Nguyen JN, Hwang CE, et al. Increased Lower Extremity Injury Risk Associated With Player Load and Distance in Collegiate Women's Soccer. *Orthopaedic Journal of Sports Medicine* 2021;9(10):23259671211048248.
34. Enright K, Green M, Hay G, et al. Workload and Injury in Professional Soccer Players: Role of Injury Tissue Type and Injury Severity. *Int J Sports Med* 2020;41(02):89-97. [published Online First: 04.12.2019]
35. Schwartz J, Spix C, Touloumi G, et al. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *Journal of Epidemiology & Community Health* 1996;50(Suppl 1):S3-11.
36. Murray NB, Gabbett TJ, Townshend AD, et al. Calculating acute: chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages. *Br J Sports Med* 2017;51(9):749-54.
37. Sedeaud A, De Larochelambert Q, Moussa I, et al. Does an Optimal Relationship Between Injury Risk and Workload Represented by the "Sweet Spot" Really Exist? An Example From Elite French Soccer Players and Pentathletes. *Frontiers in Physiology* 2020;11:1034.
38. Almon S. The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society* 1965:178-96.
39. Armstrong B. Models for the relationship between ambient temperature and daily mortality. *Epidemiology* 2006:624-31.
40. Elliott P, Shaddick G, Wakefield JC, et al. Long-term associations of outdoor air pollution with mortality in Great Britain. *Thorax* 2007;62(12):1088-94.
41. Abrahamowicz M, Beauchamp ME, Sylvestre MP. Comparison of alternative models for linking drug exposure with adverse effects. *Statistics in medicine* 2012;31(11-12):1014-30.
42. Sylvestre MP, Abrahamowicz M. Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in medicine* 2009;28(27):3437-53.
43. Rogalski B, Dawson B, Heasman J, et al. Training and game loads and injury risk in elite Australian footballers. *Journal of science and medicine in sport* 2013;16(6):499-503.
44. Cousins BE, Morris JG, Sunderland C, et al. Match and Training Load Exposure and Time-Loss Incidence in Elite Rugby Union Players. *Frontiers in Physiology* 2019;10:1413.
45. Ryan MR, Napier C, Greenwood D, et al. Comparison of different measures to monitor week-to-week changes in training load in high school runners. *International Journal of Sports Science & Coaching* 2021;16(2):370-79.
46. Ramskov D, Rasmussen S, Sørensen H, et al. Interactions between running volume and running pace on injury occurrence in recreational runners: A secondary analysis. *Journal of Athletic Training* 2021 doi: 10.4085/1062-6050-0165.21
47. Carey DL, Blanch P, Ong K-L, et al. Training loads and injury risk in Australian football—differing acute: chronic workload ratios influence match injury risk. *British journal of sports medicine* 2017;51(16):1215-20.
48. Eckard TG, Padua DA, Hearn DW, et al. The relationship between training load and injury in athletes: a systematic review. *Sports medicine* 2018;48(8):1929-61.
49. Lolli L, Batterham AM, Hawkins R, et al. Mathematical coupling causes spurious correlation within the conventional acute-to-chronic workload ratio calculations. *British Journal of Sports Medicine* 2019;53(15):921-22. doi: 10.1136/bjsports-2017-098110
50. Griffin A, Kenny IC, Comyns TM, et al. The Association Between the Acute: Chronic Workload Ratio and Injury and its Application in Team Sports: A Systematic Review. *Sports Medicine* 2019:1-20.
51. Zouhal H, Boullosa D, Ramirez-Campillo R, et al. Acute: Chronic Workload Ratio: Is There Scientific Evidence? *Frontiers in Physiology* 2021;12

52. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in medicine* 2019;38(11):2074-102.
53. Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic biology* 2004;53(5):793-808.
54. Anderson D, Burnham K. Aic myths and misunderstandings 2006 [Available from: https://sites.warnercnr.colostate.edu/anderson/wp-content/uploads/sites/26/2016/11/AIC-Myths-and-Misunderstandings.pdf accessed 2022-02-22.
55. Glatting G, Kletting P, Reske SN, et al. Choosing the optimal fit function: comparison of the Akaike information criterion and the F-test. *Medical physics* 2007;34(11):4285-92.
56. Bjørndal CT, Bache-Mathiesen LK, Gjesdal S, et al. An Examination of Training Load, Match Activities, and Health Problems in Norwegian Youth Elite Handball Players Over One Competitive Season. *Frontiers in Sports and Active Living* 2021;3(36) doi: 10.3389/fspor.2021.635103
57. Borg DN, Nguyen R, Tierney NJ. Missing Data: Current Practice in Football Research and Recommendations for Improvement. *Science and Medicine in Football* 2021:1-6. doi: 10.1080/24733938.2021.1922739
58. Moons KG, Donders RA, Stijnen T, et al. Using the outcome for imputation of missing predictor values was preferred. *Journal of clinical epidemiology* 2006;59(10):1092-101.
59. Peters SA, Bots ML, den Ruijter HM, et al. Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. *Journal of clinical epidemiology* 2012;65(6):686-95.
60. Van Buuren S. Flexible imputation of missing data: Chapman and Hall/CRC 2018.
61. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Scandinavian journal of medicine & science in sports* 2006;16(2):83-92.
62. Nielsen RO, Bertelsen ML, Ramskov D, et al. Time-to-event analysis for sports injury research part 1: time-varying exposures. *Br J Sports Med* 2019;53(1):61-68.
63. Miles A. Obtaining predictions from models fit to multiply imputed data. *Sociological Methods & Research* 2016;45(1):175-85.
64. Team RC. R: A language and environment for statistical computing.: Foundation for Statistical Computing, Vienna, Austria, 2021.
65. Bache-Mathiesen LK. Modelling the cumulative effect of training load on injury risk: R code repository 2022 [1.0.0:[Available from: https://github.com/lenakba/modelling-training-load.
66. Sylvestre MP, Evans T, MacKenzie T, et al. PermAlgo: Permutational algorith to generate event times conditional on a covariate matrix including time-dependent covariates. *R package version 11* 2015
67. slider: Sliding Window Functions 0.2.1 [program]: The Comprehensive R Archive Network, 2021.
68. Zeileis A, Grothendieck G. zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software* 2005;14(6):1–27. doi: 10.18637/jss.v014.i06
69. Ulrich J. TTR: Technical Trading Rules. *R package version 0242* 2020
70. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. *Journal of statistical software* 2011;43(8):1.
71. Association GAotWM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists* 2014;81(3):14-18.
72. Datatilsynet. The anonymisation of personal data: Norwegian Data Protection Agency; 2017 [Available from: https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/anonymisation/ accessed 2021-10-14.