Lena Kristin Bache-Mathiesen // Improving the Methodology of Training Load and Injury Risk Research

DISSERTATION FROM THE
NORWEGIAN SCHOOL OF
SPORT SCIENCES
2023

Lena Kristin Bache-Mathiesen

Improving the Methodology of Training Load and Injury Risk Research:

An Analysis of Analyses

NORWEGIAN SCHOOL
OF SPORT SCIENCES

nih.no

Lena Kristin Bache-Mathiesen

Improving the Methodology of TrainingLoad and Injury Risk Research:

An Analysis of Analyses

DISSERTATION FROM THE NORWEGIAN SCHOOL OF SPORT SCIENCES • 2023

ISBN 978-82-502-0610-6

Table of contents

Acknowledgements

Everyone says taking a PhD is like riding a rollercoaster. If that's the case, whoever engineered my rollercoaster experimented with new loops and corkscrews. Thankfully, a host of conductors kept an eye out and saw me through to the end. Supervisors, co-authors, coworkers, family and friends. And a dog. Without the support of everyone at the Oslo Sports Trauma Research Centre, the Department of Sports Medicine at the Norwegian School of Sport Sciences, and Aspetar Orthopaedic and Sports Medicine Hospital, I wouldn't have finished this PhD. At least not in one piece.

First of all, thank you to Morten Wang Fagerland, my main supervisor. Morten championed this project through the trying times of the Corona virus pandemic. No matter what was going on, he made sure to set aside time for my statistical questions. Questions where I overanalyzed and complicated simple matters, and Morten grounded me with simple, elegant solutions. For that, I am ever grateful.

Considerable thanks also to supervisor Thor Einar Andersen. I didn't (don't) always understand clinical theory, and this is where Thor Einar plucked whatever abstract thoughts floated in my mind and tied them to clinical practice. He taught me how to write about complex statistical methods in articles targeted towards clinical researchers. Our findings would not have reached the community and made the same impact without his contribution.

Alongside my supervisors stood the knowledge and support of several co-authors. Thank you, first, to Torstein Dalen-Lorentsen. Torstein was exceedingly patient with me and my questions about training load and football—I'm sure I was not the only one at Norwegian School of Sports Sciences who didn't know how many players are on a football team. His dissertation was a valuable steppingstone for this project, and he was among the first to question the methodological practices which this project addresses.

Many thanks to Benjamin Clarsen. He was among those who identified the need for improved methodology in training load and proposed this PhD project, and so, without his passion, this project may never have been. His insight into sports physiology was invaluable in our work. Thank you, also, for correcting my Americanized British English mistakes.

I also had the privilege of collaborating with scientists at the Research Department of Aspetar Orthopaedic and Sports Medicine Hospital. Thank you Montassar Tabben for helping me with the Qatar Stars League database, answering all my questions about the data, and joining me in thought-provoking scientific discussions. Thanks also Karim Chamari for a warm welcome and tour of the world-class sports medicine hospital, which really made me feel at home in the ASPREV office. Without the contribution of Montassar Tabben and Karim Chamari, we wouldn't have seen the final set piece of this thesis.

I would like to offer my gratitude to everyone in the ASPREV team, and at Aspetar Orthopaedic and Sports Medicine Hospital in general; together they ensured my stay in Qatar was fun and memorable. Many thanks to Roald Bahr and Marco Cardinale for arranging the internship, and ensuring the transfer went as smooth as possible. Thanks to Bahar Hassanmirzaei, who offered a warm welcome and friendly working environment during my stay in Qatar. Thanks also to Rodney Whiteley and Christopher Skazalski. They did not only teach me a great deal about biomechanics and volleyball, they also shared their experiences of moving to Qatar. That way, I could learn from their mistakes, and do them anyway. Special thanks also to Neshi Arif and Anna Kochergina for expedient untying of administrative knots, which allowed the timely completion of this research.

I wish to thank all members of the Oslo Sports Trauma Research Center and the Department of Sports Medicine for a friendly and supportive working environment. Thanks to Tone Helga Rasmussen Øritsland and Solveig Sunde for efficient administration. Thanks also to Sigmund Alfred Anderssen, Jostein Steene-Johannessen and Roald Bahr for excellent leadership and motivating us even through the challenges of the Corona virus pandemic. I also send thanks to my fellow PhD candidates (or previous, as some have now graduated as doctors!) for sharing experiences as a PhD candidates and making work days a lot more fun.

During this period, I also had the pleasure of collaborating with Christian Thue Bjørndal and Siv Gjesdal at the Department of Sport and Social Sciences, and Live Luteberget at the Department of Physical Performance. From the other side of the rivers that separate our departments, they welcomed me, and I realized those rivers were, in reality, narrow creaks that I could step over at any time. Thanks to Christine Holm Moseid and Grethe Myklebust for joining me on this excursion.

Special thanks, also, to Christian Thue Bjørndal for sharing the Norwegian elite youth handball data, and for following up on all my questions (nagging, really) on the project. And, to Garth Theron, for helping me with the Norwegian Premier League database at late hours and answering all my questions by e-mail.

I am grateful to everyone that paved the way for me to take this PhD. Thank you to all statisticians in Health West for providing a great learning environment. Special thanks to Ingvild Dalen, who held my hand through the field of statistics and improved my understanding of causal inference; and to Karl Ove Hufthammer, for mentoring me in R statistical programming. Your knowledge on the topic is invaluable and I would not have reached this level without your guidance. Thanks also to Anne-Marie Fenstad for notifying me about the PhD project application. I also wish to thank Anna Runemark and Richard Ian Bailey, my M.Sc. supervisors at the University of Oslo, who ignited my curiosity and provided me the confidence to consider a career in research.

Lastly on this plentiful list, I dedicate the best space to thank my friends and family. Thank you, Kim Bache-Mathiesen for respecting my busiest periods, for arranging off leash days for our dogs (so that I may work in peace), and for inviting me to cozy afternoons. Thank you Mette Bertnes, for supporting me throughout these three years by always picking up the phone, whenever that may be (morning and night), whatever the topic (easy and difficult), and wherever the place (Norway and Qatar). Also thank you for insisting that I shouldn't work on vacation days—even though I didn't always listen, I appreciate the effort. I also offer my gratitude to Dag Bache-Mathiesen for improving my sanity with morning walks in the neighborhood, for helpful career and travel advice, and support.

Thank you to Marit Elise Ødegaard, Lars Brendmoe, Siri Brendmoe, and Trine Brendmoe for dog-sitting, for hospitality, for helping out with errands on tough days, and for cozy times. I needed your fun and engaging stories, even the crazy ones, to ground me from all the science.

Finally, I thank Marius Brendmoe. Thank you so much for your caring support. Without it, I would've decayed to an office zombie long ago.

Thank you all for a wonderful PhD-coaster.

Lena Kristin Bache-Mathiesen

Oslo, November 2022

List of papers

This thesis is based on the following original research papers (available in the Appendices), which are referred to in the text by their Roman numerals:

I. Bache-Mathiesen, L., Andersen, T. E., Clarsen, B., Fagerland, M. W. (2021). Handling and reporting missing data in training load and injury risk research. *Science and Medicine in Football.* doi.org/10.1080/24733938.2021.1998587

II. Bache-Mathiesen, L., Andersen, T. E., Dalen-Lorentsen, T., Clarsen, B., Fagerland, M. W. (2021). Not straightforward: modelling non-linearity in training load and injury research. *BMJ Open Sport & Exercise Medicine*, *7*, e001119. doi.org/10.1136/bmjsem-2021-001119

III. Bache-Mathiesen, L., Andersen, T. E., Dalen-Lorentsen, T., Clarsen, B., Fagerland, M. W. (2022). Assessing the cumulative effect of long-term training load on the risk of injury in team sports. *BMJ Open Sport & Exercise Medicine*, *8*, e001342. doi.org/10.1136/bmjsem-2022-001342

IV. Bache-Mathiesen, L., Andersen, T. E., Dalen-Lorentsen, T., Tabben, M., Chamari, K., Clarsen, B., Fagerland, M. W. (2022). A new methodological approach to training load and injury risk: separating the acute from the chronic load [Manuscript submitted for publication]. Department of sports medicine, Norwegian School of Sport Sciences.

Abbreviations

ACWR	Acute:Chronic Workload Ratio
AIC	Akaike's Information Criterion
AW	Average Width of confidence intervals
EWMA	Exponentially Weighted Moving Average
CART	Classification And Regression Tree
CI	Confidence Interval
DLNM	Distributed Lag Non-Linear Model
FP	Fractional Polynomial
GPS	Global Positioning Systems
GDPR	General Data Protection Regulation
HR	Hazard Ratio
MAR	Missing At Random
MCAR	Missing Completely At Random
MCSE	Monte Carlo Standard Error
MI	Multiple Imputation
MNAR	Missing Not At Random
NIH	Norwegian School of Sport Sciences
OR	Odds Ratio
P	P-value
PB	Percent Bias
PMM	Predicted Mean Matching
QSL	Qatar Stars League
RA	Rolling Average
RB	Raw Bias
RCS	Restricted Cubic Splines
REDI	Robust Exponential Decreasing Index
RMSE	Root-Mean-Squared Error
RPE	Rating of Perceived Exertion
SD	Standard Deviation
SE	Standard Error
SI	Single Imputation
sRPE	session Rating of Perceived Exertion

Summary

Background Sport injuries burden professional and recreational athletes. In 2021, Norwegian hospitals operated 1 462 anterior cruciate ligaments, and 62% of these happened during sports activity. To prevent injuries, it may be possible to change the training load. Unfortunately, how training load can be altered to achieve desired outcomes is unknown, because the relationship between training load and injury risk has proven difficult to study. The ability of currently used statistical methods to capture this complex relationship is either limited, or unknown. Consequently, studies have employed a plethora of statistical approaches. Systematic reviews have reported inconsistent and even conflicting findings both within and between studies, and declared the studies too variable to compare in analyses. Experts have questioned the evidence supporting training load as an injury prevention tool, and called for improved statistical methodology. Despite this, few studies have recommended alternatives, and those who have, have not tested the methods' accuracy or precision. The validity of recommended methods is therefore unknown. To improve research on injury prevention programs, knowledge is needed on how to statistically determine the relationship of training load and injury risk.

Aims To identify statistical methods suitable for assessing the relationship between training load and injury risk. Specifically, to find methods for dealing with 1) missing data, 2) non-linearity, 3) time-dependent effects, and 4) the effects of relative training load.

Main Methods We analyzed three football datasets and one handball dataset: Norwegian Premier League men's football (42 players, 38 injuries), Norwegian U-19 football (81 players, 81 injuries), Norwegian elite youth handball (205 players, 471 injuries), and Qatar Stars League (QSL) football (1 465 players, 1 977 injuries). In all Norwegian cohorts, training load was defined as the number of minutes in training/match activity multiplied by the athlete's rating of perceived exertion on a scale from 1 to 10 (sRPE). The Norwegian Premier League data additionally had measures of distance and speed registered by Global Positioning Systems (GPS) devices in football. In the QSL cohort, training load was defined as the number of minutes in football training/activity.

The Norwegian Premier League football and Norwegian elite youth football were the basis for three simulation studies (*Paper I–III*). We simulated a relationship between training load and probability of injury under different scenarios of missing data, non-linearity, and time-dependent effects. With the aid of accuracy and uncertainty measures, we compared the ability of various statistical methods to model the simulated relationships in the respective scenarios.

Regression analyses were used to check whether there were any signs of non-linearity between sRPE and injury risk in the three Norwegian cohorts (*Paper II*), and also signs of time-dependent effects between training load and injury risk in the handball and QSL cohorts (*Paper III–IV*). In addition, we applied a novel approach of estimating the effect of recent training load relative to past training load on injury risk (relative training load) on the Norwegian elite U-19 and QSL data (*Paper IV*).

Main Results In each of the simulations, the performance of a few methods stood out from the rest. Firstly, for handling missing data, multiple imputation using predicted mean matching had, generally, the lowest percentage bias of all compared methods, and had acceptable bias (< |5%|) up to 50% missing data in sRPE and up to 90% missing data in the total distance GPS measure. Secondly, when we modelled parabolic non-linear relationships, fractional polynomials, quadratic regression and restricted cubic splines had the lowest root-mean-squared error, and highest coverage of 95% prediction intervals. Lastly, in the simulation of time-dependent effects, the distributed lag non-linear model was the only method that accurately modelled more than one scenario. It had the lowest root-mean-squared error and the narrowest 95% confidence intervals, by far, compared with the other methods.

The handball model presented a parabolic J-shaped relationship between sRPE and injury risk ($p < 0.001$). The QSL model displayed time-dependent effects, where effect estimates of past training load decreased exponentially for each day in the past. The QSL model also showed highest injury risk at low levels of past training load, lowest risk at medium levels, and intermediate risk at high levels of past training load, for each level of recent training load. This demonstrated that relative training load can be modelled with this novel approach.

Conclusion Missing data in training load should be imputed with multiple imputation using predicted mean matching. Researchers in training load and injury risk should consider the potential for non-linearity and time-dependent effects, and explore such effects by specifying fractional polynomials or restricted cubic splines in distributed lag non-linear models. Modelling recent and past training load separately can be used to study the effects of relative training load on injury risk.

Sammendrag på norsk (Summary in Norwegian)

Bakgrunn Idrettsskader er en byrde på idrettsutøvere og mosjonister. Norske sykehus opererte 1462 korsbåndskader i 2021, og 62% av disse var takket være deltakelse i idrett. For å forebygge skader, kan det være mulig å endre treningsbelastningen. Dessverre er det ukjent hvordan man kan endre treningsbelastning for å oppnå effekt, fordi sammenhengen mellom treningsbelastning og skade har vært utfordrende å forske på. Hvorvidt vanlige statistiske metoder kan fange opp denne sammenhengen er enten begrenset, eller ukjent. Som konsekvens, har studier i feltet brukt mange forskjellige statistiske tilnærminger. Systematiske oversikter rapporterer inkonsistente og til og med konflikterende resultater både i og på tvers av studier, og har erklært studiene for ulike til å sammenligne i analyser. Eksperter har stilt spørsmål til evidensen som står bak nådagens anbefalinger til treningsbelastning som verktøy for skadeforebygging, og etterspurt forbedret statistisk metodologi. Til tross for dette, har få studier foreslått alternativ, og de som har, har ikke testet metodenes nøyaktighet eller presisjon. Validiteten til de anbefalte metodene er derfor ukjent. For å forbedre forskning på skadeforebygging, kreves mer kunnskap om hvordan man skal statistisk fastslå sammenhengen mellom treningsbelastning og skaderisiko.

Formål Identifisere statistiske metoder som egner seg for å studere sammenhengen mellom treningsbelastning og skaderisiko. Spesifikt, finne metoder for å håndtere 1) manglende data, 2) ikke-linearitet, 3) tidsavhengige effekter, og 4) effekten av relativ treningsbelastning.

Hovedmetoder Vi gjorde analyser på tre fotballdatasett og et håndballdatasett: Herrefotball i eliteserien (42 spillere, 38 skader), norsk under-19 fotball (81 spillere, 81 skader), unge norske håndballspillere fra fem idrettsgymnas (205 spillere, 471 skader), og «Qatar Stars League» (QSL) fotball (1 465 spillere, 1 977 skader). I alle norske kohorter var treningsbelastning definert ved antall minutter i trening/kamp aktivitet, ganget med utøverens vurdering av intensiteten på en skala fra 1 til 10 (sRPE). Dataene fra eliteserien hadde i tillegg målinger på distansen løpt i fotball, registret med globalt posisjonssystem (GPS). I QSL kohorten var treningsbelastning definert som antall minutter med fotball trening/kamp-aktivitet.

Dataene fra eliteserien og under-19 fotball var utgangspunktet for tre simuleringsstudier (*Artikkel I–III*). Vi simulerte en sammenheng mellom treningsbelastning og sannsynlighet for skade under ulike omstendigheter med manglende data, ikke-linearitet, og tidsavhengige effekter. Ved hjelp av målinger på nøyaktighet og usikkerhet, sammenlignet vi evnen til ulike statistiske metoder til å modellere den simulerte sammenhengen.

Regresjonsanalyser ble brukt til å sjekke om det var noen tegn til ikke-linearitet mellom sRPE og skaderisiko i de tre norske idrettskohortene (*Artikkel II*), i tillegg til tegn til tidsavhengige effekter mellom treningsbelastning og skaderisiko i håndball og QSL kohorten (*Artikkel III-IV*). Til sist utforsket vi en ny tilnærming til å modellere effekten av nåtidstreningsbelastning relativt til fortidstreningsbelastning (relativ treningsbelastning) på skaderisiko i norsk under-19 og QSL fotball kohortene (*Artikkel IV*).

Hovedfunn Noen få metoder skilte seg ut fra de andre i simuleringene. For å håndtere manglende data, hadde multippel imputering generelt den laveste prosent skjevhet og hadde akseptabel skjevhet ($< |5\%|$) t.o.m. 50% manglende data i sRPE-målinger, og t.o.m. 90% manglende data i distansen løpt (GPS-måling). Da vi modellerte parabolske ikke-lineare sammenhenger, hadde fraktale polynomer, annengradspolynomer og kubiske spliner lavest kvadratisk gjennomsnittsfeil, og høyest dekning av 95% prediksjonsintervaller. Til sist, i simuleringen av tidsavhengige effekter, var «distributed lag non-linear models» den eneste metoden som modellerte mer enn et scenario med tilstrekkelig nøyaktighet. Den hadde lavest kvadratisk gjennomsnittsfeil og de smaleste 95% konfidensintervaller, med stor margin sammenlignet med de andre statistiske metodene.

Håndballmodellen formet en J-formet sammenheng mellom sRPE og skaderisiko ($p < 0.001$). QSL-modellen viste tidsavhengige effekter, hvor effektestimatet til fortidsstreningsbelastning ble eksponentielt mindre for hver dag tilbake i tid. QSL-modellen indikerte også høyest skaderisiko ved lave mengder fortidstreningsbelastning, lavest skaderisiko ved moderate mengder treningsbelastning, og intermediær risiko for høye mengder treningsbelastning, for hvert nivå av nåtidstreningsbelastning. Dette demonstrerte at relativ treningsbelastning kan modelleres med denne nye tilnærmingen.

Konklusjon Manglende data i treningsbelastningsmålinger bør imputeres med multippel imputasjon. Forskere i feltet for treningsbelastning og idrettsskader burde vurdere potensiale for ikke-linearitet og tidsavhengige effekter, og utforske disse ved å spesifisere fraktale polynomer eller kubiske spliner i «distributed lag non-linear models». Nåtidsstreningsbelastning og fortidstreningsbelastning kan modelleres separat for å studere effekten av relativ treningsbelastning på skaderisiko.

1 Introduction

Sports injuries impact the sports industry substantially, as they can hamper athlete and team performance in all sports (Hoffman et al., 2020; Hägglund et al., 2013; Williams et al., 2016). Injuries can lead to player absence (Hoffman et al., 2020), player retirement (Okholm Kryger et al., 2015), prolonged, chronic pain (Myklebust et al., 2003), and chronic conditions such as osteoarthritis (Myklebust & Bahr, 2005).

With the ultimate goal of injury prevention, researchers in sports medicine science and sports science strive to identify risk factors for injury—in particular, modifiable factors (Bahr & Krosshaug, 2005; Cameron, 2010; van Mechelen et al., 1992). One potential, modifiable risk factor is training load: The mechanical, physiological and psychological load the athlete has been exposed to over a period of time (Windt & Gabbett, 2017). This can be expressed by the intensity, duration, frequency and/or pattern of training and/or competition activities that subject the athlete to exertion (Bourdon et al., 2017). The terms "training load", "load", and "workload" are used interchangeably in the literature (Ide et al., 2021), and hereafter, I will use the term "training load" to capture the same theoretical construct.

In 2014, a new method for analyzing training load and injury risk in sport was introduced in a study of cricket (Hulin et al., 2014). The approach was developed further and presented more formally in 2016 (Blanch & Gabbett, 2016), and since its introduction, the number of studies assessing training load and sports injuries increased substantially (Figure 1; Gabbett, 2018). The majority of these studies claimed an association between training load and injury (Eckard et al., 2018; Griffin et al., 2020), and training load interventions were recommended (Gabbett et al., 2016). Consequently, training load monitoring and management strategies gained traction as preventative measures for injury (Akenhead & Nassis, 2016; Bourdon et al., 2017; Gabbett et al., 2016). Experts raised concerns, however, about the evidence supporting training load management strategies to mitigate injuries (Gamble, 2013; Franco M. Impellizzeri et al., 2020a; Windt et al., 2018), and the methodological approaches were under particular scrutiny (Windt et al., 2018).

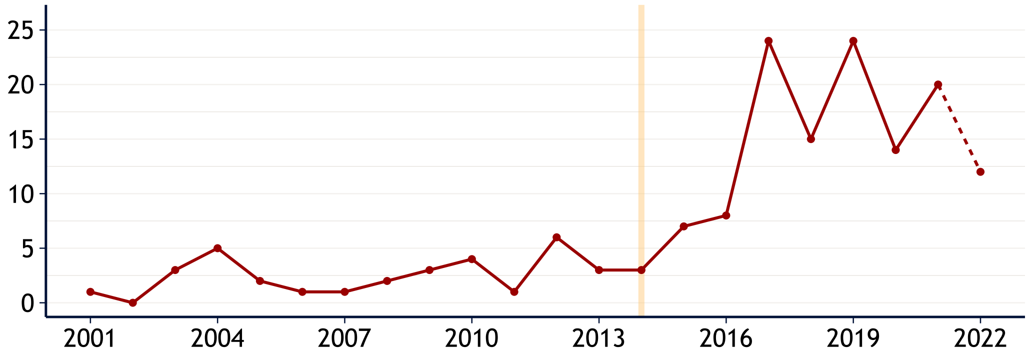

Figure 1. The yearly number of publications on training load and sports injury risk increased markedly after new statistical approaches were introduced in Hulin et al. (2014) (vertical yellow line). Calculated on Bache-Mathiesen (2022c), updated November 1st 2022 (n = 159). Since 2022 had yet to pass, dotted line represents the uncertainty of 2022 data. A similar trend was found in Gabbett (2018).

The methodological discussion at the time was two-fold. One part concerned the frequent use of suboptimal statistical methods (Windt et al., 2018). The other concerned an ongoing discussion on how to model the potential effect of training load on the risk of sports injury, while handling the multitude of assumptions established in its complex etiology (Bittencourt et al., 2016; Meeuwisse et al., 2007; Windt & Gabbett, 2017). Currently employed statistical methods were considered inadequate in meeting these assumptions (Franco M Impellizzeri et al., 2020; Menaspà, 2017). When assumptions of statistical approaches are violated, they are likely to lead to biased results, which in turn, leads to incorrect conclusions. A randomized controlled trial conducted in 2018 found no effect of a training load management intervention on health problems in 394 elite youth football players (Dalen-Lorentsen, Bjørneboe, et al., 2021), although the intervention had previously shown promising results from the aforementioned observational studies (Gabbett, 2016). Gross statistical errors were later uncovered in these observational studies (F. Impellizzeri et al., 2019). Injury prevention interventions based on faulty conclusions might burden coaches and athletes with ineffective measures, or worse, increase injury risks, as speculated by Gamble (2013) on training load management implemented in cricket.

Despite these concerns, few studies have ascertained how the training load and injury risk relationship should be statistically modelled. Most studies have explained the limitations of currently employed statistical methodology without concrete recommendations for alternatives (Lolli et al., 2018; Menaspà, 2017; C. Wang et al., 2020). Although the field of sports medicine has some general methodology guidelines for injury research, they are not tailored specifically for the context of training load (Nielsen, Shrier, et al., 2020; Ruddy et al., 2019). A handful of studies have provided recommendations to meet specific assumptions (Nielsen et al., 2019; Williams, West, et al., 2017). No study has considered the problem of modelling training load and injury risk as a whole. Therefore, how to handle one assumption without violating another—handling them all collectively—is unclear. This has led to statements such as in Toresdahl et al. (2022): "Lastly, [method used] has limitations that have been previously described […] another approach may be more appropriate for runners training for a marathon for which additional research is needed." Many researchers are aware of the issues of currently employed statistical methodology, but have nowhere else to turn for practical, evidence-based guidelines (more examples are Johansson et al., 2021; Nakaoka et al., 2021; Wang et al., 2022).

How training load affects injury, and how it can be used to prevent injury, will remain unknown unless suitable methods are employed. Further understanding on how to model this complex relationship is therefore needed to provide researchers with the tools necessary to reach these goals. The overall aim of this dissertation was to identify and recommend suitable methods for research on the relationship between training load and injury risk.

1.1 Background: The training load-injury etiology

1.1.1 The causal pathway from training load to injury

In 1994, Meeuwisse introduced a theoretical model where athletes are susceptible to injury through a combination of multiple risk factors (Meeuwisse, 1994). Internal risk factors are descriptions of the athlete (age, fitness, injury history), while external risk factors are extrinsic exposures such as equipment, weather, and playing field conditions. He argued that during a sports activity, the pre-disposed athlete, as determined by internal risk factors, experiences not only external risk factors, but factors which are directly associated with the onset of the injury—the inciting injury event (Figure 2). While the internal and external risk factors affect injury probabilistically, the inciting event is described by Meeuwisse (1994) as the necessary cause of injury (Figure 2). Bahr and Holme (2003) built on Meeuwisse's model with the idea that the inciting event can be distant in time from the outcome (the injury), especially for gradual onset (overuse) injuries. Gradual onset injuries are characterized by stiffness, pain and other symptoms which may periodically occur, subside, worsen, and re-occur (Finch & Cook, 2014). Past injuries, gradual and sudden onset, may render the athlete susceptible to future injuries. Therefore, in 2007, Meeuwisse et al. integrated the concept of recursive injuries to the multifactorial injury model.

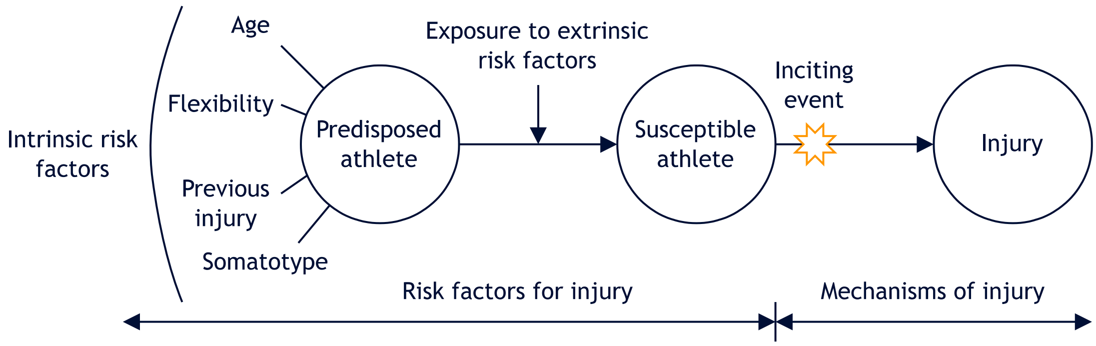

Figure 2. The multifactorial model of injury, adapted from Meeuwisse (1994) Figure 2.

Training load is a potential external risk factor. It is a multidimensional construct (Bourdon et al., 2017), and has traditionally been delineated into two components: external and internal load (F. M. Impellizzeri et al., 2019). External load describes the physical exertion which an athlete has performed, such as the distance run, number of jumps performed, or the duration of the training session (Bourdon et al., 2017). Internal load describes the athlete's physiological and psychological response to the activity (F. M. Impellizzeri et al., 2019).

In Kalkhoven et al. (2021), external load is defined as a surrogate measure of the mechanical load. The mechanical load is the physical force and pressure, caused by the external load, that strains and potentially damages the tissue (Vanrenterghem et al., 2017). Repeated mechanical load incites a biomechanical response of remodeling and repair; the tissue builds adaptations for improved resilience and performance in the future (Verheul et al., 2020). However, if the accumulation of strain and damage exceeds the capacity for the tissue to repair and adapt, the tissue may develop maladaptations or weaken instead, thus becoming more susceptible to injury (Verheul et al., 2020). Such a biomechanical process may also be instigated or exacerbated by the physiological load (Vanrenterghem et al., 2017). Exposure to an inciting event causes a mechanical load in excess of that tolerated under normal circumstances, or reduces the tolerance levels to a level which a normal mechanical load cannot tolerate, resulting in an injury (Bahr & Krosshaug, 2005).

Banister et al. (1975) hypothesized that training stimuli causes aftereffects that can both positively and negatively affect performance, a theory known as the fitness–fatigue model. Gabbett (2016) and Windt and Gabbett (2017) adapted the fitness–fatigue model to the context of injury risk. They argued that athletes only sustain sports injuries when theys participate in activities that expose them to training load. Every time they perform an activity, they risk experiencing the fatiguing effects of training, as well as potential maladaptations from overtraining (Figure 3). On the other hand, though, adequate training loads are necessary to build beneficial physiological adaptations such as high aerobic capacity and strength, which are associated with decreased injury risk. How training load may both increase and decrease injury risk is known as the training–injury paradox (Gabbett, 2016).

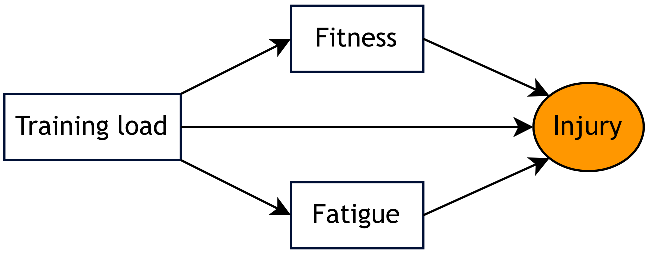

Figure 3. A visualization of the training load and injury paradox outlined in Gabbett (2016) and Windt et al. (2017). Training load has a direct effect on injury through sheer exposure. Training load affects injury indirectly through the path of fitness, which decreases risk, and through fatigue, which increases risk.

Blanch and Gabbett (2016) formed a model of assessing training load and injury risk based on the training–injury paradox. The past training load in the previous (commonly 3–4) weeks, denoted the "chronic" load, is considered to reduce injury risk through building fitness, as opposed to the "acute" current training load, often measured as the latest week of training, which is thought to increase injury risk. A sudden "spike" in training load exposure, that is, a large amount of training load in the current week (acute load) compared with previous weeks (chronic load), is thought to increase risk, as the tissue is not prepared, i.e. has not built the necessary fitness, to tolerate the oncoming training load (Gabbett, 2016). This is sometimes referred to as the "too much, too soon"-theory (Franco M. Impellizzeri et al., 2020b; Soligard et al., 2016). Given these assumptions, the relative training load—the amount of training load incurred recently relative to that incurred in the past—may also be important determining injury risk (Blanch & Gabbett, 2016).

1.1.2 Complexity in the training load-injury etiology

Traditionally, injury risk studies have approached risk factor identification from a reductionist point of view (Ruddy et al., 2019). In such a paradigm, the whole risk factor–injury risk system is considered an additive sum of its parts. Humans are, however, a non-linear system (Fonseca et al., 2020). In contrast to additive systems, relationships between variables are not constant in non-linear systems; they change with the state of the system (Stern et al., 2021). Lich et al. (2013) called for a shift from a reductionist approach to a complex system's approach. Risk factors for sports injuries may interact multiplicatively, have non-linear relationships with the risk of injury, and form negative feedback loops (diminishing returns, Bittencourt et al., 2016). The contribution of each risk factor towards the risk of injury may also be different in different sports. Collectively, Bittencourt et al. (2016) called this framework the sport-specific "web of determinants" (Figure 4). Addressing such complexity require changes in the study design as well as the statistical methodology (Bekker, 2019).

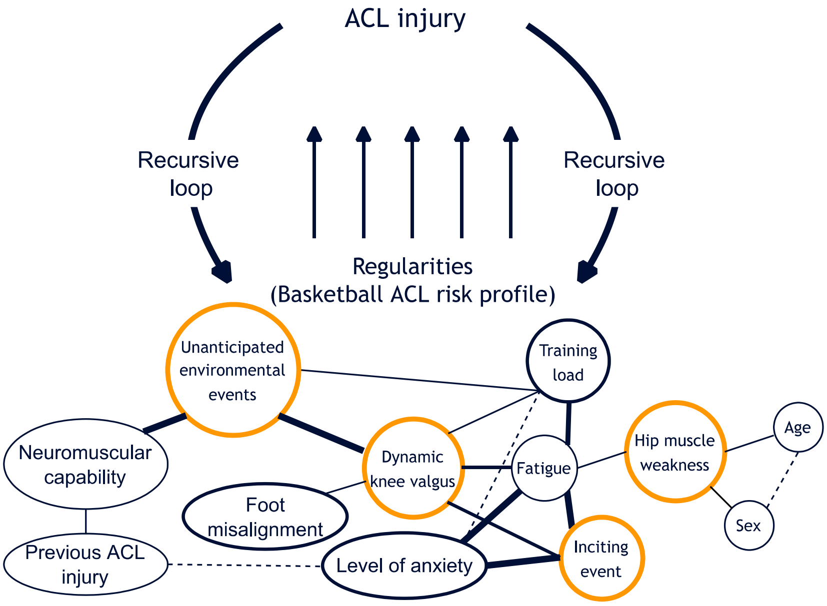

Figure 4. The web of determinants for anterior cruciate ligament (ACL) injury in basketball, adapted from Bittencourt et al. (2016) Figure 2A. The variables at the bottom contribute to injury risk with different weights. Thicker frames indicate variables with strong effects on ACL injury risk. Dotted lines are weak interactions, strong lines are strong interactions. The combination of risk factors, and how they interact, may be different for different sports.

The training load-injury etiology outlined previously indicates that such complexities are present in how training load affects injury risk. A property of non-linear systems is that risk factors of injury can have different and sometimes opposite effects during different states (Stern et al., 2020; Stern et al., 2021). Training load appears to be an example of this: If the tissue state is strained, training load can increase risk, but if the tissue state is properly prepared, training load can further build fitness and decrease risk—that is, training load has a direct effect in some states and an inverse effect in other states. This corresponds to a non-linear relationship between training load and injury. Gamble (2013) hypothesized that this non-linear relationship might be U-shaped; both low and high levels of training load increase risk, and moderate training load levels reduce risk. Furthermore, Blanch and Gabbett (2016) assumed that relative training load and injury risk forms a J-shaped relationship (hockey stick curve), though the claim was later criticized for having neither foundation in clinical rationale nor in statistics (F. Impellizzeri et al., 2019).

Interestingly, the (non-linear) direction of effects may be *time-dependent*. In the "too much, too soon"-theory, acute training load increases risk, while chronic training load reduces risk (Gabbett, 2016). Renfree et al. (2021) described how a brief, high-volume activity on a single day may substantially increase risk relative to the same volume of activity spread across several days. Training load may also have a *time-lagged* effect (Bhaskaran et al., 2013)—the training load on the previous day contributes to the injury risk on the following day. This effect is only indirect (Figure 3). Thus the training load on the previous day is likely less important in its contribution towards injury risk than that performed on the current day (Williams, West, et al., 2017). However, not only the previous day affects injury risk, but also the day before the previous day, which may be of even less importance (Williams, West, et al., 2017). We can assume that this pattern of effect continues the further we go back in time. This is known as a protracted time-lagged effect (Gasparrini, 2014; Richardson, 2009). The injury risk at any given time is the result of multiple training load exposure events of different intensities sustained in the past. Epidemiologists call this an exposure-lag-response relationship (Gasparrini, 2014).

In summary, the hypotheses discussed so far imply the following assumptions:

- The relationship between training load and probability of injury may be non-linear.
- The current effect of training load may be a cumulative sum of effects resultant of training load exposures in the past.
- The size and direction of effect may depend on time since the training load exposure.
- The effect of training load may depend on the size of recent exposure relative to the size of past exposure (relative training load).

In this dissertation, these four assumptions were central in the development of methods for analyzing training load and injury risk.

1.2 Project overview

1.2.1 Studying training load and injury risk: causal inference or prediction?

The most appropriate approach for analyzing training load and injury risk depends, ultimately, on the research question (Shrier et al., 2022). Where one statistical model would be suitable to answer which risk factors contribute alongside training load to injury risk, another would be more applicable to answer how training load increases longevity of sports-participation until an injury occurs. Therefore, a single approach will never cover the entirety of training load–injury risk research. One major perspective to consider is whether the study aims for causal inference or prediction (Nielsen, Simonsen, et al., 2020). Although some modeling methods can be used for both, the model requirements, and how to interpret the results, varies between the two (Shmueli, 2010), and some methods are ideal for one, but not the other (Pepe et al., 2004).

Bittencourt et al. (2016) called for a paradigm shift from risk factor identification to injury prediction. They argued that to prevent injuries, we must first be able to predict injuries. Bittencourt et al. further expressed how methods for prediction (here, machine learning) can handle the complex systems etiology of injury risk and lessen the amount of assumptions (Bittencourt et al., 2016). Methodology in prediction studies differ markedly from those of causal inference (Shmueli, 2010). Although there are exceptions, prediction studies are more interested in developing a parsimonious prediction model than assessing effect sizes of a single predictor. The contents of the prediction model are less important; the focus is on the combined predictive ability of the model. Variable selection is done by regularization methods, and weak predictors are discarded (Zumeta-Olaskoaga et al., 2021). In causal inference, the exposure(s) of interest is always included in the analysis. Stratification and adjustment strategies are employed, among other options, to approach an unbiased estimate of the effect of the exposure of interest on the outcome (Shrier & Platt, 2008; Stovitz & Shrier, 2019). Other variables may be analyzed to assess how much of the effect is explained through different causal pathways, but the effect sizes usually do not determine their model inclusion (unlike in studies of prediction).

By virtue of assessing a single, modifiable exposure of interest, studies on training load and injury risk are more geared towards causal inference than prediction. In addition, despite ambiguous aims of assessing a "relationship" or "association" (Hulin & Gabbett, 2019; Nielsen, Bertelsen, et al., 2020), the studies often recommend interventions (for example Shaw et al., 2021), and sometimes consider confounding, both of which implies an aim of causal inference (Hernán, 2018). The end goal is to determine how training load can be modified to change injury outcomes. We therefore chose to approach this project from a causal approach, that is, when considering methods, explaining how dimensions of training load affects injury risk was prioritized over predictive modelling.

1.2.2 The van Mechelen sequence of sports injury prevention

The sequence of sports injury prevention (Figure 5), developed by van Mechelen et al. (1992), has been used by sports scientists for systematic injury prevention. The first step in the sequence is to map the current frequency and severity of an injury problem. The second step is to find causes of injury and develop injury etiologies. The third step is to develop injury prevention measures based on the knowledge gathered in step 2 and introduce them to the target population. The last step is to repeat step 1 to determine whether the implementation improved the injury problem in question. Observational training load and injury risk studies aiming for causal inference or identifying populations at risk are in step 2 in the van Mechelen sequence. This thesis project is hence about providing a statistical toolbox for conducting such studies (Figure 5).

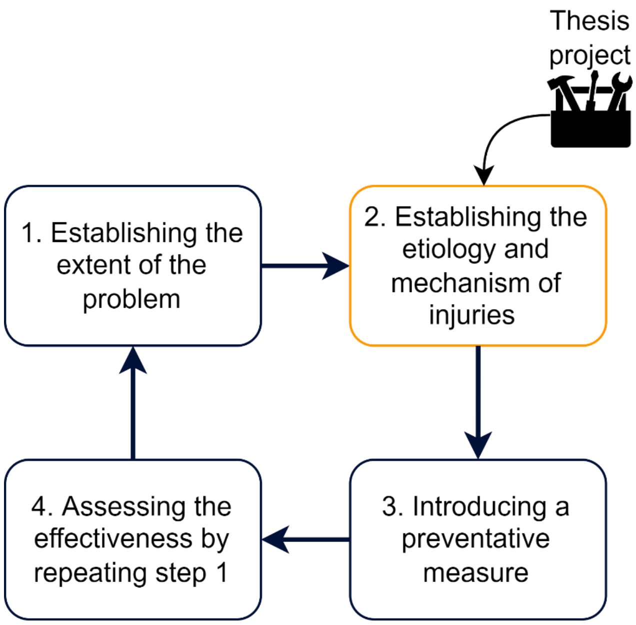

Figure 5. Training load and injury risk studies currently belong in step 2 in the sequence of injury prevention (van Mechelen et al., 1992). This dissertation aimed to provide statistical tools for researchers conducting training load and injury studies in step 2.

1.2.3 The traditional sequence of training load data preparation and analysis

To meet the aims of step 2 in the van Mechelen sequence, training load and injury risk studies generally follow the same process of measuring, preparing and analyzing training load (Figure 6). First, training load is measured longitudinally in the field by either external, internal, or both dimensions of training load measures (Windt et al., 2018). The raw measures are processed and prepared for the next step by cleaning errors, handling data quality concerns, and deriving compound measures. In step 2, analysts aggregate the measures by time intervals to study effects of long-term training load (A. Wang et al., 2021). If the study aims to assess relative change in training load, the aggregated values are further processed in the optional step 3. In step 4, also optional, the modified training load values are categorized (Figure 6, Dalen-Lorentsen, Andersen, et al., 2021). This was done in 82% of 138 papers published between 2001 and 2021 (calculated on Bache-Mathiesen, 2022c). Finally, in step 5, the relationship between the processed measures and injury measures are analyzed with methods such as hypothesis testing and/or regression (Windt et al., 2018). In summary, there are currently 4 steps of data preparation of training load measures before analysis, with multiple choices at each step (Figure 6).

Researchers often include multiple variations of training load measures and injury definitions – analyzing different measures, time intervals, aggregation methods, and calculation choices in the same study (Miguel et al., 2021; Udby et al., 2020). The results are often inconsistent or conflicting (a few examples are: Sedeaud et al., 2020; Toresdahl et al., 2022; West et al., 2020), making them difficult to interpret (Franco M. Impellizzeri et al., 2020b). Reviews of the training load and injury risk field have reported difficulties in comparing studies due to between-study-variation (Maupin et al., 2020; Sniffen et al., 2022), and the potential for p-hacking has also been a central concern (Dalen-Lorentsen, Andersen, et al., 2021; Franco M. Impellizzeri et al., 2020b). A consensus on statistical methodology is needed to solve these issues. In this dissertation, we have focused on understanding which methods are suitable under which scenarios to develop concrete recommendations, and, we have considered methods that require fewer subjective choices if possible.

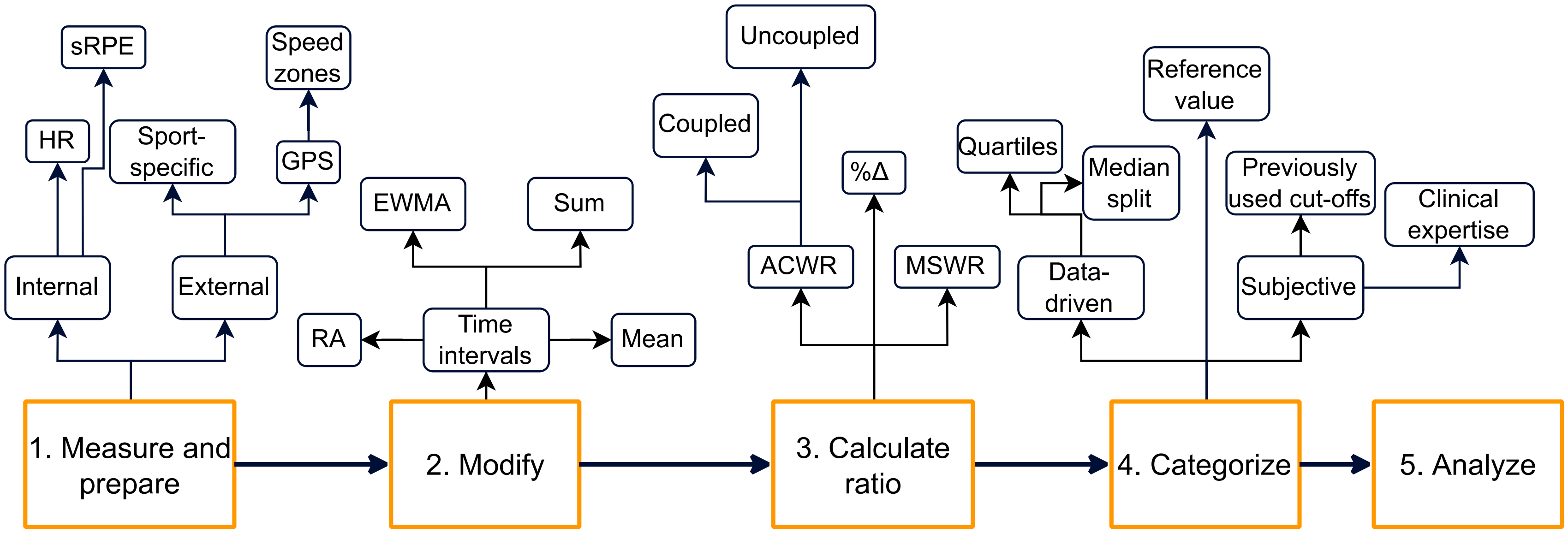

Figure 6. Flowchart of the current training load data process before analysis of the relationship between training load and injury risk. A tree of methodological options at each step is drawn to illustrate the number of choices taken before the final step. Only a limited number of options are shown. The steps are as follows: (1) Training load is measured longitudinally in the field, and prepared by handling missing data and other data quality concerns. (2) Prepared training load measures are partitioned into time intervals and aggregated. (3) If relative training load is of interest, a ratio is calculated on the modified observations. (4) The continuous absolute or relative training load values are categorized into groups. (5) The measure(s) processed through steps 1–4 are analyzed with hypothesis testing or regression to determine their relationship with injury risk. ACWR, Acute:Chronic Workload Ratio; EWMA, Exponentially Weighted Moving Average; GPS, Global Positioning Systems; HR, Heart Rate; MSWR, Monotony and Strain Workload Ratio; RA, Rolling Average; sRPE, session Rating of perceived exertion.

The scientific articles that comprise this thesis project can be organized by the steps that they address in the training load and injury risk data preparation process (Figure 7).

- *Paper I* deals with missing data in training load measures and only addresses step 1
- *Paper II* compares methods of modification, ratios, categorization and analysis in handling non-linearity, and therefore touches on steps 2–5
- *Paper III* compares methods of modification, ratios, and analysis in determining the cumulative effect of training load on injury risk, and as such, addresses steps 2, 3, and 5
- *Paper IV* considers how to analyze relative training load without using a ratio, and addresses only step 5.

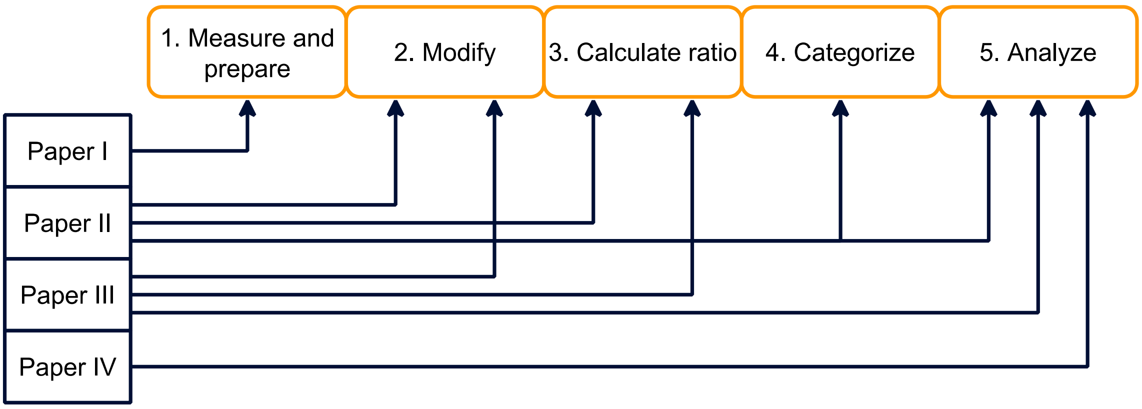

Figure 7. The four scientific articles in this thesis project, organized by which step in the training load data preparation process they address.

1.3 Analyzing training load and injury risk: from measure to model

1.3.1 Measuring training load (step 1)

External training load measures include, but are not limited to, time in activity, the distance run, the number of training sessions completed, participation in matches or competition events; sport-specific measures such as the height of jumps in volleyball, or the number of balls thrown in baseball (Mehta, 2019; Miguel et al., 2021; Skazalski et al., 2018; Udby et al., 2020). To measure internal load, researchers can choose between physiological load metrics like heart rate and oxygen consumption (F. M. Impellizzeri et al., 2019; Mallo & Dellal, 2012; Owen et al., 2015), and psychological load such as athlete-reported intensity (Borg et al., 1987). A consensus statement recommended assessing both external and internal load in studies on training load and injury risk (Bourdon et al., 2017), which is most commonly done in football according to a recent review (Miguel et al., 2021). This recommendation has recently been challenged: Different dimensions of training load—external, mechanical, physiological, and psychological—may have different causal pathways to injury (Kalkhoven et al., 2021; Vanrenterghem et al., 2017), and separate external load constructs—intensity, frequency, duration—may contribute differently to tissue adaptation and subsequent injury risk (Staunton et al., 2021). The choice among measures of external and internal training load depends on the aims and scope, and available resources, of each training load and injury risk study. The clinical rationale for why and how the training load measure is related to the injury type in question, should be the main justification for choosing one measure over another (Kalkhoven et al., 2021). Studies in causal inference may also consider prioritizing readily modifiable training load dimensions that are prime targets for intervention (Suzuki et al., 2020). To be valid, external training load measures should reflect the mediation of internal training load on injury risk (Impellizzeri et al., 2022), otherwise, modifying the measure will not lead to change in injury risk.

External load

To measure external training load, metrics from microdevices with Global Positioning Systems (GPS) technology are the most popular (Benson et al., 2020; Griffin et al., 2020). The device is securely strapped onto an athlete before activity, and estimates the running direction, speed and distance through GPS signaling. The accuracy of GPS devices depends on the signal sampling frequency, measured in herz (Hz). The devices are available commercially and scientists may choose between different manufacturers (Varley et al., 2012). While relatively expensive, they can automatically capture whole training sessions and events, and are valid and reliable for distance measures (Benson et al., 2020). The device reports the total distance run, the distance run at moderate speed, high speed and/or sprint speed (Udby et al., 2020). There is no consensus, however, on the definition of moderate, high and sprint speeds (Rago et al., 2020). In addition, GPS devices are not suitable for all sports. Cricket, volleyball and golf are a few examples of sports where training load is not predominantly from running distance. Sport-specific measures such as the number of balls bowled for cricket (Saw et al., 2011), or jump-detecting devices for volleyball and basketball (Benson, Owoeye, et al., 2021; Skazalski et al., 2018), may be more relevant measures of training load that are connected to injury risk.

Introduction

Internal load

The most popular measuring tool for internal training load is the session Rating of Perceived Exertion (sRPE), introduced by Foster et al. (2001). sRPE was used in all studies on internal training load reviewed by Griffin et al. (2020), encompassing 72% of studies in the review, and recommended in a consensus statement (Bourdon et al., 2017). First, the athlete's perceived intensity, known as the Rating of Perceived Exertion (RPE, Borg et al., 1987) is self-reported on the modified scale from 1 to 10 (Foster et al., 2001). A value of 1 is minimum intensity "Very, very easy" and 10 is maximum intensity "Maximal", while 0 denotes rest / no training load. To derive the sRPE, the RPE is multiplied by the duration of the training load activity in minutes. Studies often analyze daily sRPE, calculating sRPE for each activity independently before summing the daily scores. This results in a scale from 0 to an upper limit in the thousands, depending on the sport and population.

sRPE is considered a measure of internal load (Bourdon et al., 2017). Recently, it has been recommended to take into account that the RPE-portion of this compound measure is a measure of psychological load, only (Kalkhoven et al., 2021). The causal pathway from psychological load to injury risk may differ from that of physiological load to injury risk, and this nuance may affect methodological considerations in studies of causal inference.

Missing data in training load

To evaluate time-varying effects of training load, training load is measured repeatedly at multiple timepoints in a longitudinal design (Nielsen et al., 2019). Such data commonly includes missing values (Powney et al., 2014; Siddique et al., 2008). Athletes may not be available to be measured at all timepoints or be lost to follow-up, standardized forms may be partially complete, and GPS-data may contain errors.

Enright et al. (2020) excluded 140 (53%) injuries from analyses due to inconsistent and/or missing data. Similarly, in a 3-season football cohort, 124 out of 154 (81%) eligible injuries were excluded due to insufficient training load data (Lolli et al., 2020). As demonstrated, missing observations in training load, unless dealt with, reduce the sample size of injuries. Injuries may be rare (Bahr & Krosshaug, 2005), and to assess the relationship between training load and injury with sufficient accuracy, the analysis requires a sufficient number of events (Riley et al., 2019; van Smeden et al., 2016). Prediction studies using machine learning are at particular risk of producing overly optimistic, overfitted models if events are scarce (Sidey-Gibbons & Sidey-Gibbons, 2019; van der Ploeg et al., 2014).

Missing training load data may also lead to the removal of athletes from the analyses. In a study of 34 tennis players, 16 (47%) were removed during the study period and were not included in the analysis (Moreno-Pérez et al., 2021), even if these players had consented and participated partially. Other studies report including participants who completed > 80% of the surveys, only (Albrecht et al., 2020; Theisen et al., 2013), or completing a full season (Fanchini et al., 2018). Such practice reduces the generalizability of the study, and may, in a worst-case scenario, introduce selection bias.

Alarmingly, Borg et al. (2022) found that only 11% of studies on football topics, or involving football players, reported whether or not they had any missing data. In the training load and injury risk field, this number was unknown, although, few studies (33%) reported how they handled missing data (Windt et al., 2018). Those who had, used varying methodology (A. Wang et al., 2021); from mean (Brink et al., 2010) to median imputation (Johnston et al., 2019), to complete case analysis (Malone et al., 2018) and linear regression imputation (Esmaeili et al., 2018). McCall et al. (2018) stated "There is currently no best practice for dealing with missing data, and averages were chosen […]". Therefore, in *Paper I*, we performed a systematic review of the literature to map current practices for reporting and handling missing data. The results from the review were used to inform our methodological choices in the subsequent study.

Missing data can be Missing Completely at Random (MCAR), meaning the probability of missing does not depend on any other factor or variable, observed or otherwise (Janssen et al., 2010). For example, blood samples that were accidentally dropped to the floor. In that case, no selection bias is introduced. However, clinical data and participant-reported data are more often Missing at Random (MAR, Barnett et al., 2017; Janssen et al., 2010), a case where the missingness is dependent on other variables collected in the study. An example would be if men were more wary of reporting their injury status than women. MAR data can be imputed by using the other variables as predictors (Janssen et al., 2010). At the very least, the analyst may discover and report the dependency alongside the analyses, which is not possible when data are Missing Not at Random (MNAR).

Under MNAR, missingness is dependent on unobserved factors (Janssen et al., 2010). As an example, suppose that—unbeknownst to the researchers—athletes who are in a certain socioeconomic status are less likely to report their injury status. Thanks to missingness, the study has unidentified selection bias. More serious are cases where missingness is dependent on the variable in which the missingness resides. One can imagine players who are uninjured are less likely to report that they are in fact, uninjured, and will cause injury frequencies to be overestimated. Or similarly, athletes who have higher training loads are too busy to respond to the survey about training load, and the analysts will receive a false picture of the average load of the cohort. Such bias is difficult, if not impossible, to detect, but are likely in training load and injury research.

Multiple imputation, by machine learning or more commonly used options like predicted mean matching, use the other variables in the dataset to predict observations in place of the missing data. Such methods have shown solid results in the field of statistics (Chhabra et al., 2017; White & Carlin, 2010), and medicine (Jakobsen et al., 2017), under both MCAR and MAR, and is considered best practice in some milieus (van Ginkel et al., 2020). To achieve valid results, multiple imputation requires a correctly specified imputation model: the model that predicts the training load observations replacing the missing data (Sterne et al., 2009). Having too few or too weak predictors in the imputation model may introduce bias. Important predictors of common training load measures are currently unknown. It is also unclear whether multiple imputation can perform under the common condition of limited information in a training load and injury risk study. For instance, in a study which has only collected external or internal training load measures and not the other, or, has not collected variables for confounder adjustment. Therefore, we addressed these knowledge gaps in *Paper I*.

Although multiple imputation methods are considered best practice, simpler methods may be suitable in certain situations. Complete case analysis, also known as listwise deletion, is the practice of deleting the rows with missing data and running the analysis on the complete cases. Given a large enough sample size, this has several advantages (Marshall et al., 2010). It cannot introduce unrealistic or impossible values, usually retains the distribution of data, and is easy to use. Unless otherwise specified, the statistical software packages SPSS, R and Stata run complete case analysis by default (IBM, 2020; Kabacoff, 2011; UCLA, 2021). Under MAR, however, it can cause selection bias (White & Carlin, 2010), and it reduces the sample size of the data, thus reducing statistical power.

A potential alternative to deletion methods is mean or median imputation. This method replaces the missing observation with the average of the observed values. While it is easy to perform, it may reduce the variability of the dataset and skew distributions (Barzi & Woodward, 2004), which may bias analyses performed on the imputed data. On the positive side, it retains all the data, preserving sample size and power. In training load and injury risk studies of small sample sizes, it may be more prudent to choose mean imputation over complete case analysis. Although mean imputation can give more biased estimates, it will protect the power by saving potentially rare injury events from deletion. Choosing between these two methods may be a matter of cost-benefit at a study by study basis. In *Paper I*, we investigated whether these simpler methods were adequate in some cases.

A. Wang et al. (2021) raised concerns with the variation in how means were calculated before mean imputation in training load and injury risk studies, and there seemed to be no consensus in how the mean should be calculated to achieve the lowest amount of bias. Benson, Stilling, et al. (2021) compared different variants of mean imputation, and multiple imputation with gradient boosted regression, for imputing training load measured by the Rating of Perceived Exertion (RPE, Borg et al., 1987) on the CR10 scale (Foster et al., 2001). Their results suggested that the performance of mean imputation was dependent on how the mean was calculated. If a basketball player was missing an observation, the mean of all other players training on the same day was more informative than the mean of all previous observations for that player. Multiple imputation had superior performance over mean imputation, though they recommended mean imputation if more advanced methods were not available.

Benson, Stilling, et al. (2021) implored future researchers to determine how the session Rating of Perceived Exertion (sRPE) should be imputed, the most common measure of training load (Griffin et al., 2020). Since sRPE is the product of two factors, the RPE on a scale from 0 to 10, and the activity duration in minutes, it is unclear whether sRPE should be calculated before, during, or after imputation. Given the additional complexity of the relationship between these variables, we addressed how to impute sRPE in *Paper I*, and also tested two variants of mean imputation to see if results in Benson, Stilling, et al. (2021) were reproducible in a different context.

Benson, Stilling, et al. (2021) only looked at 1% missing. They also gauged performance by comparing the imputed data with the observed data using root-mean-squared error, which Van Buuren (2018) cautioned against doing in isolation. The purpose of imputation is to retain the observed data on other variables, so that all the observed data is used in the analysis of interest, i.e. a regression model. Ideally, performance is assessed by the amount of bias and uncertainty introduced to the analysis through the method used to impute data (Van Buuren, 2018, chapter 2.5). Van Buuren (2018), chapter 2.6, demonstrated that a method with reasonable root-mean-squared error when comparing imputed versus observed data may still be biased in regression modelling. In *Paper I*, we assessed performance of imputation by comparing the accuracy of a logistic regression model run on data imputed with different approaches.

1.3.2 Modification of training load measures (step 2)

Injury cannot occur unless athletes participate in a training or sporting activity which exposes them to external training load (Windt & Gabbett, 2017). Most likely, sports injury frequency will increase with time spent in activity. Studies which explore the difference between players with high levels of training load and those with low levels of training load (Dennis et al., 2003), or explore the relationship between training load during a week and injuries in the same week (Hulin et al., 2014, Murray et al., 2017, Bowen et al., 2017), have discovered a positive association through the effects of direct exposure (Figure 3, C. Wang et al., 2020). They may also discover an inverse association—that decreased training load increases injury risk (e.g. Moreno-Perez et al., 2021). This pattern arises when athletes cease training due to injury, and have periodically lower training loads (Carey et al., 2017). Training load management for injury prevention is aimed at improving fitness and reducing fatigue—these are the indirect causal paths between training load and injury risk (Figure 3). To understand the training load and injury picture in a way useful for developing injury prevention tools, the effects of past, both short and long-term training load need to be assessed.

The protracted, time-dependent properties of training load is, however, challenging to take into account (Nielsen et al., 2019; Windt & Gabbett, 2017). To meet the assumption that past training load cumulatively effects injury risk, studies have parted the training load data into time intervals of equal length that typically span one or more calendar weeks (Step 2 in Figure 6), known as weekly blocks or windows (Mandorino, Figueiredo, Condello, et al., 2022; Ryan et al., 2021). Observations spanning one or more time windows are then aggregated to capture past training load (Mandorino, Figueiredo, Condello, et al., 2022; Udby et al., 2020).

Traditionally, the aggregations, such as the weekly mean training load, move iteratively in a sliding window from one week to the next (A. Wang et al., 2021). This is an inefficient use of the data; six days of injury observations are skipped for each interval. Furthermore, it fails to capture nuances in training load changes, such as a week including a recovery day, and a week without (Menaspà, 2017). More grievously, in studies which aggregate weekly training loads, athletes who are injured early in the week and taken out of practice can cause the illusion that low training load amounts increase injury risk (C. Wang et al., 2020). Carey et al. (2017) proposed moving the aggregations iteratively in a sliding window from one day to the next to combat these issues.

Partitioning training load by calendar weeks has also been considered an arbitrary time-delineator for many sports (Franco M Impellizzeri et al., 2020). In team sports, coaches often periodize recovery and training according to matches, in so-called micro-cycles (Malone et al., 2015). For example, in football, a micro-cycle consists of recovery days after the previous match, the training days before the next match, and the next match (Figure 8). A match is denoted M. Given k, the number of days, negative values of k indicate a day of training: M-1 is the training day before a match, M-2 two days before a match; each micro-cycle includes M-1, M-2, …, M-k training days before the next match. Recovery days are denoted with a positive k: M+1 is the recovery day after a match, M+2 is the recovery day two days after a match. Some calendar weeks may have multiple matches, and thus stretch over multiple micro-cycles (Figure 8). In other cases, more than a week may pass between matches. Under this assumption, training schedules depend not on calendar days, but on match schedules. Coyne et al. (2022) suggested adjusting training load time windows to micro- or meso-cycles of training. A sliding window of aggregation from one micro-cycle to the next, and from one day to the next, were both used in *Paper II*.

To aggregate the time intervals of training load data, studies have employed different statistical approaches (Udby et al., 2020; A. Wang et al., 2021; Windt et al., 2018). The most frequently used method is the rolling average (A. Wang et al., 2021). Researchers have, however, identified considerable disadvantages of using rolling averages to deal with time-lagged effects (Gasparrini, 2016; Menaspà, 2017). The method assumes that all training load exposures in the past, plus the exposure on the current day, contribute equally to injury risk.

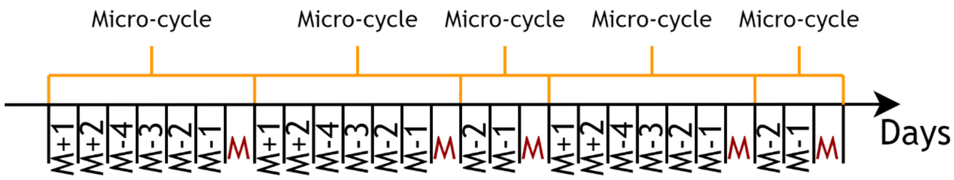

Figure 8. Illustration of a football micro-cycle. Each micro-cycle period consists of all activity before a new match (M). That is, recovery days after the previous match as well as the training days before the next match. Days denoted with negative numbers are training days before the next match (M-1; being the day before the match, M-2; two days before a match, and so on). Days with positive numbers are recovery and training days after a match (M+1; being the day after a match, M+2; two days after a match). The number of days between matches varies by the match schedule. How a team plan their training and recovery activities varies, and is dependent on the teams' philosophy.

Introduction

The exponentially weighted moving average (EWMA) assumes that training load exposure further back in time affects injury risk less than observations closer in time (Williams, West, et al., 2017). EWMA has been used to aggregate chronic load in the calculation of relative training load (Dalen-Lorentsen, Andersen, et al., 2021; Hamlin et al., 2019; Nakaoka et al., 2021). Many studies calculating EWMA do so in addition to rolling averages—analyzing both (Arazi et al., 2020; Enright et al., 2020; Nakaoka et al., 2021; West et al., 2020; Xiao et al., 2021). Based on more instances of statistically significant results ($p < 0.05$) the EWMA has been considered a more sensitive measure of detecting injury risk (Murray et al., 2017; S. West et al., 2021; West et al., 2020). Since the advent of EWMA in 2017, studies have nevertheless continued to use rolling averages (Albrecht et al., 2020; Hildebrandt et al., 2020; Malone et al., 2020; Moreno-Pérez et al., 2021).

In a recent commentary, EWMA was considered insufficient to meet the assumptions of training load outlined previously (C. Wang et al., 2020). It cannot be calculated for time-intervals with missing training load observations, which are common in longitudinal data (Jeličić et al., 2009). In addition, unlike the rolling average, it cannot be calculated on incomplete time windows. To calculate a 4-week EWMA, the researcher must discard the first 27 days of training load and injury data before calculation of the first EWMA value. Finally, the difference between the weights at Day 28 and Day 27 increases as the decay constant gets closer to zero (C. Wang et al., 2020). C. Wang et al. (2020) specify: "The contribution of the load on Day 100 is 1.9 times the contribution of the most recent load [Day 0] to the weighted average, even though the most recent load should contribute the most weight."

The Robust Exponential Decreasing Index (REDI, Moussa et al., 2019) was proposed as an alternative to the EWMA. REDI is a weighted rolling average which specifies a replacement value for missing observations. Thus, REDI can be calculated from training load data with missing observations and incomplete time windows, and it had improved performance over EWMA in a training load and injury risk study (Sedeaud et al., 2020). The methodological study did not, however, compare REDI with EWMA in the instance of no missing data (Moussa et al., 2019), which is a likely scenario if researchers impute missing data before analysis (Hecksteden et al., 2022).

Most studies that have so far considered the performance of RA, EWMA and REDI have discussed theoretical rationale, practicality and mathematics in editorials and commentaries (Menaspà, 2017; C. Wang et al., 2020). Only a handful of studies have compared training load values after calculation (Moussa et al., 2019), and assessed model fit on observed data (Sedeaud et al., 2020; West et al., 2020). So far, the ability of these methods to model the cumulative effect of long-term training load on injury risk has not been assessed in a study where the true relationship is known. We therefore compared these methods in a simulation in *Paper III*.

Although training load as an exposure has special properties, the protracted time-lagged effects are akin to those analyzed in environmental epidemiology. In this field of research, scientists assess the effects of long-term exposures such as background radiation, pollution, temperature and humidity, on outcomes such as number of hospitalizations and cancer occurrence (Bhaskaran et al., 2013). Here, the exposures, like training load, have a long-term, likely small-to-moderate and a cumulative effect on the outcome. To ascertain such effects, scientists use distributed lag models (Bhaskaran et al., 2013), a method first developed in econometrics (Almon, 1965). In the last few decades these models have been extended to handle non-linearity (Armstrong, 2006), and generalized for application on all types of longitudinal data (Gasparrini, 2014). These are called Distributed Lag Non-Linear models (DLNM, Gasparrini, 2011). We determined whether the DLNM can be applied in a sports science setting in *Paper III*.

1.3.3 Ratio calculation for relative training load (step 3)

Both the absolute training load (i.e. distance run) and the relative training load (i.e. distance run this week relative to distance run previous week) are thought to have an effect on injury risk (Gabbett, 2016; Tysoe et al., 2020). The training load and injury paradox outlined previously in section 1.1.1, has emphasized the need to capture both the protective effects of training load, and the detrimental effects, through appropriate measures. This was the aim of the Acute:Chronic Workload Ratio (ACWR, Blanch & Gabbett, 2016). It is, traditionally, the sum of the training load sustained the last seven days (acute period) divided by the rolling average of the last twenty-eight days (chronic period, Lolli et al., 2019), though the method of calculation and the length of the acute and chronic periods can vary at the analyst's discretion (Dalen-Lorentsen, Andersen, et al., 2021). When the acute load (current week) is lower than the chronic load (previous weeks), the ACWR is lower than one, and the athlete is considered to be prepared for the training load in the current week, and injury risk is, in theory, reduced (Gabbett, 2016). When ACWR is greater than one, the athlete is considered unprepared for the current demands and injury risk is increased (Gabbett, 2016).

Introduction

Since its introduction in cricket in 2014 (Hulin et al., 2014), and following refinement (Blanch & Gabbett, 2016), the ACWR became the most popular method for assessing training load (Eckard et al., 2018; Udby et al., 2020). It has, since then, been critiqued extensively (Carbone et al., 2022; F. Impellizzeri et al., 2020; Franco M Impellizzeri et al., 2020; Lolli et al., 2018; C. Wang et al., 2020; Zouhal et al., 2021). The concerns were:

- The method was invented for cricket, and may not necessarily be applicable to other sports (Franco M Impellizzeri et al., 2020).
- The method is applied to external and internal load measures with the same approach, although these dimensions are markedly different (Franco M Impellizzeri et al., 2020).
- The 1-week and 4-week time windows for the acute and chronic loads, respectively, are not sport-specific and may be arbitrary (Carey et al., 2017; S. West et al., 2021).
- A complete time window must pass before first calculation, reducing sample size (Moussa et al., 2019).
- The number of subjective choices in time windows and calculations may tempt researchers to tinker towards desired results (Dalen-Lorentsen, Andersen, et al., 2021).
- Including the numerator in the denominator is not an accurate depiction of change in training load (C. Wang et al., 2020).
- Effect sizes appear larger due to rescaling of the metric (Lolli et al., 2018).
- The ACWR cannot be calculated on time windows with missing training load observations (Moussa et al., 2019).
- The ACWR cannot account for tapering (Franco M Impellizzeri et al., 2020): the practice of undergoing a period without training before a competition event (Mujika et al., 2004).
- The ACWR frequently fails to adjust the numerator to the denominator (Franco M Impellizzeri et al., 2020), a fundamental assumption of ratios (Curran-Everett, 2013).
- Because the ACWR is a measure of acute load, adjusted for the chronic load, it is less ideal for studies that are more interested in assessing the effect of chronic load (Franco M Impellizzeri et al., 2020).
- Due to normalization failure, studies often delete high ACWR values in periods following reduced training (Franco M Impellizzeri et al., 2020), such as injury recovery weeks or vacation weeks—introducing more missing data.
- Once a high-risk ACWR is observed, assuming that it has a causal effect on injury, it may be too late to manipulate the training load for injury prevention purposes (Franco M Impellizzeri et al., 2020).
- The original results were not reproduced in an observational study (Sedeaud et al., 2020), and the recommended load management intervention through the use of ACWR was debunked in an RCT (Dalen-Lorentsen, Bjørneboe, et al., 2021).

In the wake of critiques against ACWR, some studies opted for other options, and either assessed only the absolute training load (Keylock et al., 2022; Lolli et al., 2020), or calculated relative training load with the week-to-week percentage difference (Enright et al., 2020; Ramskov et al., 2021; Ryan et al., 2021). Whether the week-to-week percentage difference is an improvement over the ACWR is unknown. Also, no study has so far simulated a relationship between relative training load and injury risk and investigated whether the ACWR can detect such a relationship. *In Paper III*, we compared ACWR, week-to-week percentage difference, and DLNM, with the aim of determining how to assess the long-term, cumulative effects of relative training load.

1.3.4 Categorization to handle non-linear risk-relationships (step 4)

In the training load and injury etiology, both too little and too much training load may increase injury risk. This alludes to a parabolic relationship between training load and injury. Gamble (2013) hypothesized the presence of a U-shaped relationship; where both low and high levels of training load increase risk, and the lowest point of risk is at moderate training load levels. Lathlean et al. (2019) used fractional polynomials and discovered a U shape, and data in Sampson et al. (2018) and Weiss et al. (2017) indicated non-linear, non-parabolic relationships between training load and injury risk. Collectively, theories and evidence suggest the relationship between training load and injury risk may be non-linear, but the exact shape is unknown.

In reviews of the training load and injury risk field, the direction of the effect of training load on injury risk varied between studies (Eckard et al., 2018; Franco M. Impellizzeri et al., 2020b). Some studies reported that high amounts of training load increased risk, while others reported that low amounts increased risk. Incidentally, methods that assume linearity of the relationship between training load and injury risk, like Pearson correlations and logistic regression, were the most frequently used in the field (Windt et al., 2018). Regardless of the true relationship shape, such methods can only describe three relationship shapes: 1) a direct relationship (increase of training load = increase in injury risk), 2) an inverse relationship (decrease of training load = increase in injury risk), and 3) no relationship. Figure 9 shows the direction of the relationship between training load and injury reported by 57 studies reviewed in (Eckard et al., 2018). The disparity in relationship directions shown in Figure 9 may be explained by the existence of an underlying, non-linear relationship modelled with methods that assume linearity. When a relationship is monotonic—Y either increases or decreases when X increases—a linear model is likely to still uncover a relationship, although inaccurately (Figure 10A). If the true relationship is non-monotonic, and Y sometimes increases and sometimes decreases when X increases, a linear model is unlikely to uncover the relationship (Figure 10B). Studies that assumed linearity may have obtained different results depending on whether the true relationship between training load and injury risk is monotonic or non-monotonic in different sports and populations. Other between-study idiosyncrasies may have caused inconsistencies. For instance, the distribution of training load data, areas of data point congestion, and sample size, may have determined whether the linear model suggested a direct or an inverse relationship shape, or no relationship at all.

Since few studies have checked the linearity assumption and used methods that account for non-linear shapes, we explored whether there was any evidence of non-linearity in different sport populations in *Paper II*.

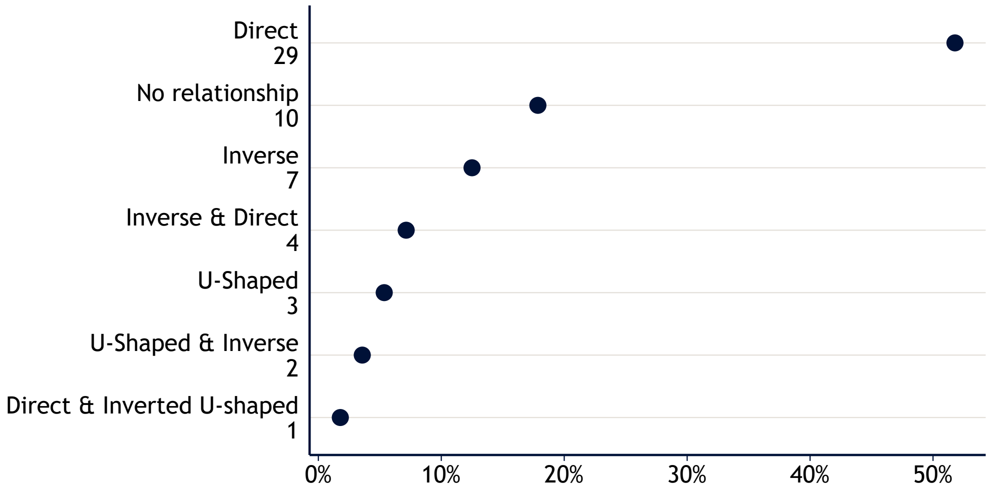

Figure 9. The direction of the relationship between load and injury risk varied considerably between the 57 studies reviewed in Eckard et al. 2017. Most reported a direct effect (51%), where injury risk increases for each increase in training load. A large portion (30%) reported inverse relationships, where injury risk decreases for each increase in training load, or U-shaped relationships, where injury risk sometimes increases and sometimes decreases with increased training load.

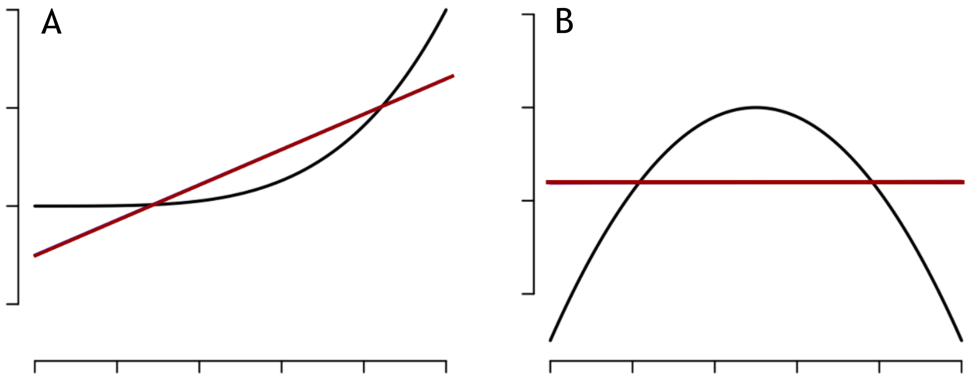

Figure 10. The discovered relationship between an independent variable (X-axis) and an outcome (Y-axis) when linearity is assumed depends on the true, underlying relationship. When the true relationship is (A) monotonic, a linear model may still discover a relationship. On the other hand, (B) a non-monotonic relationship modelled with a linear model may find no relationship at all.

Some studies reported U-shaped or non-linear results (Figure 9), which could not have been discovered with methods that assume linearity. A shared trait between many of these studies was the discretization of continuous training load variables into categories (Colby et al., 2017; Cross et al., 2016; Dennis et al., 2003; Malone et al., 2018; S. Malone et al., 2017), a method known as categorization. Categorization is often the final step before performing regression analyses or hypothesis testing (Figure 6; Dalen-Lorentsen, Andersen, et al., 2021). The practice of categorization is frowned upon in general (Frøslie et al., 2010). This method strongly assumes that the relationship between training load and injury risk remains flat within categories (Collins et al., 2016), which may be a less reasonable assumption than linearity. It also assumes that any change in risk happens at the threshold from one category to another. The user must subjectively decide cut-off values for the respective thresholds. Critics suggested such methods may encourage p-hacking: the practice of performing multiple analyses to search for significant p-values, increasing Type I error rates (Dalen-Lorentsen, Andersen, et al., 2021; Franco M. Impellizzeri et al., 2020b). A narrative review of methodology in sport injury research recommended fine-tuning categories based on predictive performance (Ruddy et al., 2019), a process which tempt p-hacking, and may lead to overfitting of prediction models and inflated optimism (Bullock et al., 2021). Such data-driven approaches also necessitate methodological choices to differ from one study to another. Consequently, reviews have struggled to perform meta-analyses due to inconsistencies in categories and reference values (Andrade et al., 2020; Griffin et al., 2020; Maupin et al., 2020; Sniffen et al., 2022). For clinicians and practitioners looking to implement best practice, the results are confusing and difficult to interpret (Franco M. Impellizzeri et al., 2020a).

In Carey et al. (2018), categorization had a poorer model fit than modeling continuous training load data, and a substantially higher rate of Type I errors (rejecting the null hypothesis when it should have been accepted) in Australian football. The authors recommended future research to check whether results are similar in other sport populations. In addition, despite raised concerns, categorization has later been recommended in methodological studies on training load and injury risk (Nielsen et al., 2019; Ruddy et al., 2019), and the number of studies using this method has not declined (Dalen-Lorentsen, 2021). Ideally, the step of categorization (Step 4 in Figure 6) is removed from the modelling process. In *Paper II*, we therefore attempted to reproduce Carey et al. (2018)'s findings in football, and assessed other ways of handling non-linearity.

A few reports of a U-shaped relationship between training load and injury risk were found in studies that employed quadratic regression (Sampson et al., 2019; Weiss et al., 2017). In some of these studies, a quadratic term was added to the regression model to test for linearity: if non-significant, it was discarded for a linear model; if significant, they categorized the load-variable to relax the linearity assumption (Ahmun et al., 2019; Sampson et al., 2018; Warren et al., 2018; West et al., 2020). This is an improvement over assuming linearity without testing, but cut-offs based on significance can be influenced by randomness. In addition, non-quadratic does not equate linear, and non-linear does not equate quadratic.

Using data from three different sports, Blanch and Gabbett (2016) modelled the relationship between relative training load (measured by ACWR) and injury risk with quadratic regression, and discovered a J-shaped relationship (hockey stick curve). In quadratic regression, training load is modelled as a continuous variable, and all the disadvantages of categorization do not apply. It is intuitive and interpretable, and may be appropriate when clinical rationale meets the model specification. Gabbett (2016) published the J-shaped figure again with highlight on the "sweet spot": the point of lowest risk. The presented figure conflicted, however, with the rationale put forward in the text. The author explained that athletes with high relative training loads (ACWR > 1) are unprepared for the demands of competition and are at increased risk of injury, whereas athletes with relative training load (ACWR < 1) have built adequate fitness and are at decreased risk. This suggests a linear relationship between relative training load and injury risk: the higher the relative load, the higher the risk. The J-shaped figure implies that, if acute load is much lower than chronic load, athletes are at increased risk of injury, which is not supported by any proposed hypotheses in the field so far. Incidentally, the J-shaped figure was based on quadratic regression. Statistical errors in developing the figure were later uncovered (F. Impellizzeri et al., 2019), and the most grievous mistake was transposing categorized relative training load data to a continuous scale for the quadratic modeling, where categories were different for the three studies included. Such a mismatch between results and theory may, however, also have been caused by the constraints of quadratic regression: it can only model a parabola, and by necessity, constrains the relationship to follow a parabola.

While the relationship between the amount of training load and injury risk has been theorized to be U-shaped (non-linear), the theories on relative training load suggest that the relationship shape may depend on the training load dimension in question (as seen in C. Wang, T. Stokes, R. Steele, et al., 2021). This may also explain some of the variation shown in Figure 9, as some studies only considered absolute, and some only relative, training load. An ideal method should be capable of uncovering both various non-linear and linear relationships. We therefore considered both relationship shapes in *Paper II*.

Quadratic regression is a subgroup of Fractional Polynomials (FP), which has been used in a single training load-injury study (Lathlean et al., 2019). Fractional Polynomials, simply put, uses polynomial transformations to estimate the association between the covariate and outcome (Royston & Altman, 1994). For researchers familiar with quadratic regression, this is intuitive and the results interpretable. Statistical simulation studies have reported that FP accurately models common non-linear relationships (Binder et al., 2013; Collins et al., 2016). On the other hand, due to the multiplicative nature of polynomials, they cannot model negative numbers nor zero.

Some studies can potentially justify adding a small constant, like 0.01, to all training load values to use FP. This is feasible if performing no training load is reasonably equivalent to performing a diminutive amount. For instance, in studies assessing the effects of past training load. In other studies, this assumption is unreasonable, or does not align with study aims (Shrier et al., 2021). Sports scientists may wish to capture biomechanical or physiological effects occurring when athletes initiate exercise. Studies on health behavior may wish to account for individuals who took the effort to go to the gym, even though they spent less than a minute at the gym. Lastly, the amount of bias introduced to the regression model would depend on the scale of the training load measure. Although adding a constant may be inconsequential for a scale from 0–3000, it would incur substantial changes on a scale of 0-10 (Foster et al., 2001).

Splines are a family of piecewise polynomials that can model local shifts in the relationship shape (Harrell Jr, 2017), such as from 0 training load to 1 in the hypothetical studies outlined above. The functions are piecewise polynomials from one interval of x, here the training load values, to the next interval. These intervals are demarcated by so-called "knots" (Gauthier et al., 2020). The order of each polynomial is the same for each interval, but the coefficients may differ. In cubic splines, the chosen polynomials are cubic. Such splines were more accurate if they were restricted to a linear shape in the tail-ends of the relationship (Stone & Koo, 1985), known as Restricted Cubic Splines (RCS). RCS has proven capable of accurately modeling non-linearity, and can model local shifts in the relationship shape (Binder et al., 2013).

The main challenge with splines is determining the number and location of the knots. Misspecified knots can bias the model. Stone (1986) commented that the number of knots was more important than the placement. When placing knots at fixed quantiles, typically, there are enough data points at each interval to inform the model sufficiently, and the risk of overly influential outliers is low (Harrell Jr, 2017). However, this may depend on the distribution of the data, and researchers may have apriori assumptions to guide knot placement. For instance, in Spanos et al. (1989), researchers aimed to predict whether a meningitis was bacterial or viral. Experiences from health professionals suggested bacterial meningitis was much more prevalent in neonatal infants than viral, but after the first year of age, their expectation was vice versa. With this information, the analysts modelled a linear spline for age of onset, and placed a knot at the age of 1. The resulting model had an area under the receiver operating curve of 0.97 (Spanos et al., 1989).

Both RCS and FP performed better than categorization in Carey et al. (2018). However, the study did not explore which method to use under which circumstances, nor whether RCS knot placement could alter results. In *Paper II*, we addressed this, and explored how to handle non-linearity in training load and injury risk research.

1.3.5 Analyzing the relationship between training load and injury risk (step 5)

A proper analysis of the relationship between training load and injury risk should be able to handle complexities in the relationship between training load and injury risk without partitioning data into time intervals, aggregating or calculating ratios before-hand, or making strong assumptions about the shape of the relationship; thus essentially skipping steps 2–4 in the data preparation process. We have previously discussed the potential for quadratic regression, splines, and fractional polynomials to handle non-linearity, and for DLNM to handle protracted time-lagged effects. These are all methods of model specification at the analysis stage.

Model specification is the process of selecting the right terms for a statistical model. This involves choosing independent variables and their functional forms, such as polynomial or logarithmic transformations. This allows the user to adapt the model to meet assumptions that are not necessarily met in the basic linear model.

One potential method of model specification is to model acute and chronic training loads separately, instead of calculating a ratio. C. Wang et al. (2020) argued that such a model would produce coefficient estimates for both acute and chronic loads, thereby allowing researchers to determine which is more important for injury risk.

The "too much, too soon"-theory postulates that the effect of acute training load on injury risk depends on the level of chronic load (Gabbett, 2016), which led to the development of the acute:chronic workload ratio. The purpose of a ratio is to adjust the numerator to the denominator, but ratios do not always succeed in doing so (Curran-Everett, 2013). Modelling the acute and chronic loads separately could potentially guarantee that the acute load is adjusted for the chronic load, and untangle the effects of acute and chronic load from each other. We therefore explored the suitability of modelling acute and chronic training loads separately to estimate both absolute and relative training load in *Paper IV*.

The "too much, too soon"-theory can also be interpreted to imply that there is an interaction between acute and chronic loads—that not just the magnitude, but also the direction and slope of the effect of acute load on injury risk depends on the level of chronic load. The presence of an interaction may cause a ratio to fail in normalizing the numerator to the denominator, and the ACWR has been criticized especially for frequently failing to normalize the acute load to the chronic load (Franco M Impellizzeri et al., 2020). In *Paper IV* we investigated whether there were any signs supporting this theory, and also modelled an interaction between acute and chronic training load in football.

As described in section 1.3.2, choosing time periods for acute and chronic loads can be challenging. The traditional ACWR aggregated training load data in weekly intervals, but it also considered the risk of injury at the weekly level. The ACWR calculation that moves iteratively from one day to the next, proposed by Carey et al. (2017), considers injury risk at the daily level. Instead of aggregating the acute loads weekly, it may be beneficial to assess the acute load at the daily level.

According to the training load-injury paradox, if players do not participate in activity on past days, they do not accrue the fitness and fatigue, which affects the risk of injury on the current day of activity (Windt & Gabbett, 2017). These are internal risk factors of the athlete (Figure 2). On the current day, training load is applied as an external risk factor, and athletes can become injured from sheer exposure to potential inciting events (Windt & Gabbett, 2017). If players do not participate in activity on the current day, they are not at risk of injury. This means that a training load observation of 0 has dramatically different effects depending on whether it was observed in the past, or on the current day. In addition, the training load planned for the current day may be more modifiable, from an injury prevention perspective, than aggregates that describe both the current day and past days, and therefore a better target for intervention (Suzuki et al., 2020). With these theories as our foundation, we determined whether there was an interaction between acute load, defined as only the current day, and chronic load, defined as all past observations in the previous four weeks, in their association with injury risk in *Paper IV*.

1.4 Aims

The overall aim of this dissertation was to determine how to assess the relationship between training load and injury risk. We endeavored to identify the most appropriate statistical methods to address specific concerns and recommend these methods for future research. Ideally, methods have high accuracy and power, and are not based on unrealistic assumptions. Each paper targeted specific sets of assumptions, and aimed to:

I. Describe the practice of reporting and handling missing data in the training load and injury risk field, and ascertain which methods introduce the least bias when handling missing data in training load measures (*Paper I*).

II. Investigate whether there is any evidence of a non-linear relationship between training load and injury risk in different sports, and determine which statistical methods are best suited to account for the assumption of non-linearity (*Paper II*).

III. Determine which statistical methods can most accurately and precisely estimate the cumulative effect of long-term training load on the risk of injury (*Paper III*).

IV. Explore the potential of modelling acute and chronic training load separately to study relative training load, and ascertain whether there is any evidence of an interaction between acute and chronic load in football (*Paper IV*).

2 Methods

2.1 Study design

In *Paper I–III*, we chose a simulation study design. To study appropriate statistical methods, it is often beneficial to use data simulation; the process of generating data that imitate real-world situations. The purpose is to infer how different methods perform under certain scenarios—such as under different probability distributions or on data of different sample sizes—by estimating and comparing power, robustness, degree of error rates and other parameters for the method(s) in question. In this dissertation, these different modeling and simulation choices were based on real-world data, to reflect the processes underlying athlete activity, so that inferences can be extrapolated for use in the respective field.

A simulation study has endless possibilities in terms of which methods to investigate, and the potential scenarios under which they may be compared. We could vary sample size, the amount of noise, the amount of missing data; the strength of the simulated relationship between training load and injury probability; add or remove dependencies between measures on the same individual—among other considerations. We therefore narrowed down the scope of each paper by limiting the number and combinations of scenarios. We aimed to mimic scenarios typical in the training load and injury risk field.

In addition to simulations, we also conducted a systematic review to map current methodological practices and guide our methodological choices in *Paper I*. We also searched for associations in an observational study design in *Paper II-IV*.

Co-authors provided statistical, clinical, and sport-specific insight to the study design. we developed a protocol before performing simulations (O'Kelly et al., 2017), available online (Bache-Mathiesen, 2021a, 2021b, 2022b). Deviations from the protocol were documented.

2.2 Participants and data

We performed one systematic review, and gained access to three European football cohorts and one handball cohort with training load and injury measures (Figure 11).

Review (*Paper I*) We performed a systematic review of the training load and injury risk literature (n = 108). Studies were extracted from the most recent, relevant reviews at the time (Andrade et al., 2020; Dalen-Lorentsen, Andersen, et al., 2021; Eckard et al., 2018; Griffin et al., 2020; Lathlean, 2017; Maupin et al., 2020; Udby et al., 2020; Windt et al., 2018), available online (Bache-Mathiesen, 2021c).

Norwegian Premier League football (*Paper I–III*) A Norwegian Premier League men's team (42 players, mean age 26 years, 38 injuries) followed through the 2019 season (Theron, 2020).

Norwegian elite U-19 football (*Paper II, IV*) A cohort of 81 players (55% male, mean age 17 years, 81 injuries) followed 16 weeks in the 2017 season (Dalen-Lorentsen, Andersen, et al., 2021).

Norwegian elite youth handball (*Paper II-III*) A cohort of 205 handball players from five Norwegian sport high schools (36% male, mean age 17 years, 471 injuries) followed through the 2018/2019 season (Bjørndal et al., 2021).

Qatar Stars League football (*Paper IV*) Eight years (2015–2022) of longitudinal data from a men's Qatar Stars League injury surveillance registry (1 465 players, 1 977 injuries).

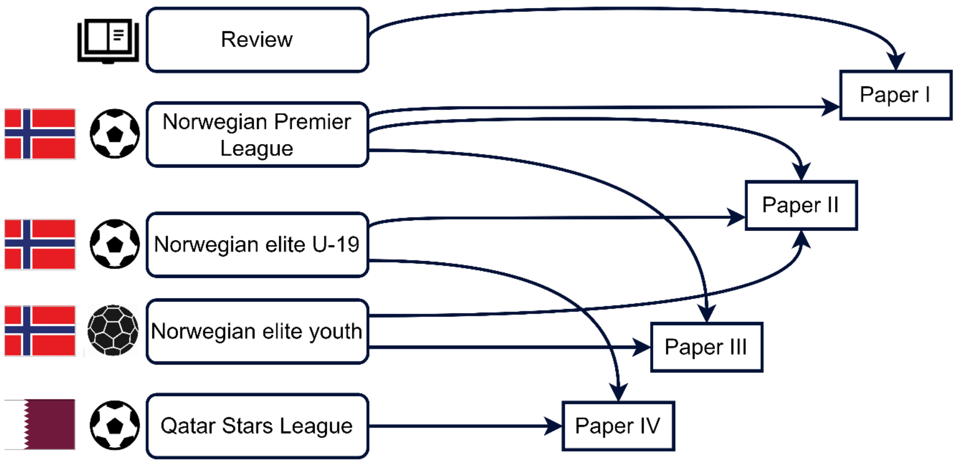

Figure 11. An illustration of which datasets were used in which papers in this dissertation.

2.3 Ethics

The Norwegian Center for Research Data approved study protocols for all Norwegian studies: Norwegian Premier League football (722773); Norwegian elite U-19 football (5487); Norwegian elite youth handball (407930). They were also approved by the Ethical Review Board of the Norwegian School of Sport Sciences (NIH). The Norwegian elite U-19 football study was additionally approved by the South-Eastern Norway Regional Committee for Medical and Health Research Ethics (2017/1015). Approvals for the Norwegian football studies are available in Dalen-Lorentsen (2021), and for Norwegian elite youth handball in Appendices, Appendix II. The Anti-Doping Lab Qatar Institutional Review Board approved the Qatar Stars League study (E2017000252), and the Aspire Zone Foundation Institutional Review Board approved a data sharing agreement signed between Aspetar Orthopaedic and Sports Medicine Hospital and the Oslo Sports Trauma Research Centre (both available in Appendices, Appendix III).

Ethical principles were followed in accordance with the Declaration of Helsinki (General Assembly of the World Medical Association, 2013; World Medical Association, 2013), except principle number 35: "Every research study involving human subjects must be registered in a publicly accessible database before recruitment of the first subject". None of these studies were registered in a publicly accessible database.

All participants, or their guardians, provided informed consent. They were assured their responses would only be available to the research team, participation was voluntary, and consent could be withdrawn at any time. Handball players were above age 15 and guardian consent was not required (Bjørndal et al., 2021). In the Norwegian Premier League football study, the on-site data collectors signed a non-disclosure agreement to prevent sharing of tactical approaches to any opponent football teams (Rønneberg, 2020).

Participant data used in this thesis was managed under the General Data Protection Regulation (GDPR, 2016/679) law of the European Parliament. Data preparation and analysis was performed on pseudoanonymized personal data stored in a secure, access-regulated server at NIH. The GDPR (Information Commissioner's Office (ICO), 2018; Privacy EU, 2018) states: "You can only use the personal data for a new purpose if either this is compatible with your original purpose, you get consent, or you have a clear obligation or function set out in law." For this reason, the Norwegian Center for Research Data determined that the Norwegian elite youth handball data had to be anonymized. In addition, for our simulation in *Paper I*, we anonymized the Norwegian Premier League data so that everything could be reproduced.

As data processor for the original aim of the Norwegian elite youth handball data (Bjørndal et al., 2021), I was also the data controller responsible for anonymizing the datasets. I followed guidelines developed by the Norwegian Data Protection Agency (Datatilsynet, 2017). First, indirectly identifiable variables were removed. This included variables such as school, playing position, BMI, etc., which, when combined, or when joined to external data, could potentially identify an individual. The player identification (ID) column was replaced with a randomly generated number with no connection to the previous identification, the original database, nor any other identifiers. The original ID column was deleted. The arrangement of individuals in the dataset were scrambled. The anonymization was performed in the command-line based statistical program R, and a new file, without a version history, was generated with the anonymized data.

After anonymization of the Norwegian elite youth handball data, I lost access to the original database. Before uploading the final, anonymous Norwegian Premier League dataset, I attempted to identify an individual through juxtaposing it with a Norwegian Premier League statistics website (Sandnes, 2021). The data passed this "motivated intruder"-test (Datatilsynet, 2017).

2.4 Training load and injury measures

The same online questionnaire was used to collect daily health status and training information from all three Norwegian sports cohorts, with Athlete Monitoring, Moncton, Canada (Norwegian Premier League) and Briteback AB online survey platform, Norrköping, Sweden (elite U-19 football, elite youth handball). Team doctors recorded corresponding information from the Qatar Stars League players in Microsoft Office Excel®, Microsoft Corporation, Readmon, WA, USA.

2.4.1 Training load definition

2.4.1.1 Session Rating of Perceived Exertion

In all Norwegian sports cohorts, the players reported the duration of each training session and match in minutes, and their perception of the intensity of the activity (psychological load) expressed as Rating of Perceived Exertion (RPE, Borg et al., 1987) on the modified Borg CR10 scale (Foster et al., 2001). On this scale, 1 is "Very, very easy" and 10 is "Maximal" intensity. The value 0 is "Rest", i.e. no activity participation. The Norwegian U-19 elite football players and elite youth handball players were prompted to respond every day, whilst the Norwegian Premier League football players were only prompted after activity completion. For these players, RPE and duration were assumed to be 0 on recovery days, that is, on the day after a match, or two days after a match (Figure 8).

For each activity, the RPE was multiplied by the duration of the activity in minutes to derive the session RPE. Daily sRPE was the sum of sRPE values per day. sRPE measures were used in all papers comprising this dissertation.

Data quality of sRPE was reported in *Paper II* Supplementary Table S1. All Norwegian datasets had missing sRPE observations: Premier League football (13%)[1], elite U-19 football (24%), elite youth handball (64%). They were imputed using multiple imputation (Madley-Dowd et al., 2019, Paper II Supplemental Figure S1) and was deemed valid (*Paper II* Supplemental Figure S2). Timeliness was considered valid in all three populations, with the mean number of days from RPE prompt to an answer at 0.01, 0.3 and 0.7 for Premier League football, elite U-19 football, and elite youth handball respectively. However, only 53% of prompts from Norwegian youth elite handball players were responded to on the same day, as opposed to 99% and 72% of the Premier League football and elite U-19 football populations.

2.4.1.2 Global Positioning Systems

In the Norwegian Premier League football study, Global Positioning Systems (GPS) were used to collect external training load measures (Rønneberg, 2020), with 10Hz sampling rate (Catapult OptimEye X4, Catapult Sports, Australia). In *Paper I*, the following GPS-variables were used: (1) total distance covered, (2) distance covered above 20 km/h (high-speed running distance), (3) distance covered above 25 km/h (sprint distance), and (4) the squared instantaneous rate of change in acceleration for three vectors of direction (*x, y,* and *z* axes) divided by 100 (player load, Boyd et al., 2011).

Daily sums were calculated for each GPS variable. Total distance was the main focus, as there was no consensus in the literature on definitions of high-speed or sprint-speed measured by GPS

[1] In *Paper II*, this was reported to be 41%. Observations that were implicitly days of no activity were erroneously considered to be missing data in this calculation. This was corrected in *Paper I*.

devices (Rago et al., 2020), and player load calculations are defined differently between GPS device manufacturers (J. J. Malone et al., 2017).

2.4.1.3 Activity duration

The Qatar Stars League registry recorded the number of minutes each football player spent in each training and/or match per day. The daily minutes in activity were a sum of all sessions on that day, imputed with multiple imputation (11% missing data*, Paper IV* Supplementary Figure S1–2).

2.4.2 Injury definition

The players in Norwegian football and handball populations reported daily whether they had experienced a new health problem. If players in the elite youth handball study reported any new health problems, they were prompted in the questionnaire to specify whether it was an injury or illness. If players in the football studies reported a new health problem, a clinician contacted them by telephone to classify it as an injury or illness in accordance with the Union of European Football Associations guidelines (Hägglund et al., 2005). Players were asked to report all physical complaints, irrespective of their consequences on sports participation or the need to seek medical attention (Bahr et al., 2020; Fuller et al., 2006).

The Qatar Stars League team doctors recorded injuries prospectively with the Sport Medicine Diagnostic Coding System classification (Orchard et al., 2020). Injuries were recorded if, due to injury, a player was unable to fully participate in training or match play (time-loss definition), and classified as either sudden or gradual onset. Validation of injuries were described in *Paper IV* Supplementary.

Only health problems classified as injuries were used in this dissertation.

2.5 Statistical analysis 1: Review

To map the current practices of handling missing data in the training load and injury field, we performed statistical analyses on the review data collected in *Paper I*. The proportion of studies reporting whether they had missing observations in the training load measure was calculated, by year and overall. The yearly percentages were plotted in a line graph to assess the trend of reporting practices.

For the studies with missing data, we determined the mean amount of missing observations in the training load variable. The percentage of studies which used each method of handling missing data was calculated.

For an estimate of study sample size, we calculated the mean, standard deviation and median number of injuries in the reviewed studies. The distribution of the number of injuries was visualized in a histogram.

2.6 Statistical analysis 2: Simulations

Stochastic simulations were performed to compare different methods of handling missing data in *Paper I* (Figure 12), non-linearity in *Paper II* (Figure 13), and cumulative protracted time-lagged effects in *Paper III* (Figure 14). More extensive detail on the methods are available in the supplementary methods files attached to the three papers respectively.

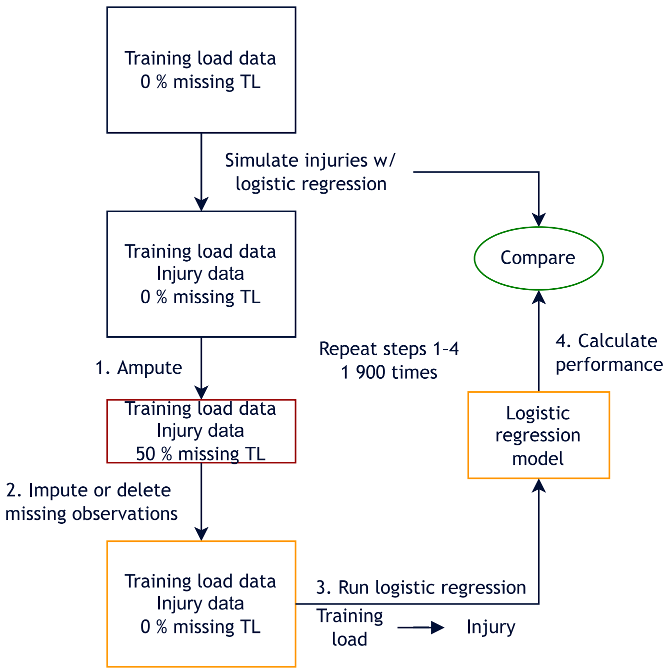

Figure 12. The simulation workflow in *Paper I*: Injuries were simulated based on probabilities from a logistic regression model. The simulation steps were then (1) introduce missing data to the training load (TL) variable, eleven different amounts under missing completely at random and three amounts under missing at random. A scenario of 50% missing is shown as an example. (2) impute or delete missing data in training load with one of five methods; (3) fit a logistic regression model with the imputed or deleted training load as the exposure and the simulated injuries as the outcome; (4) calculate performance measures and compare predicted probabilities with the simulated probabilities.

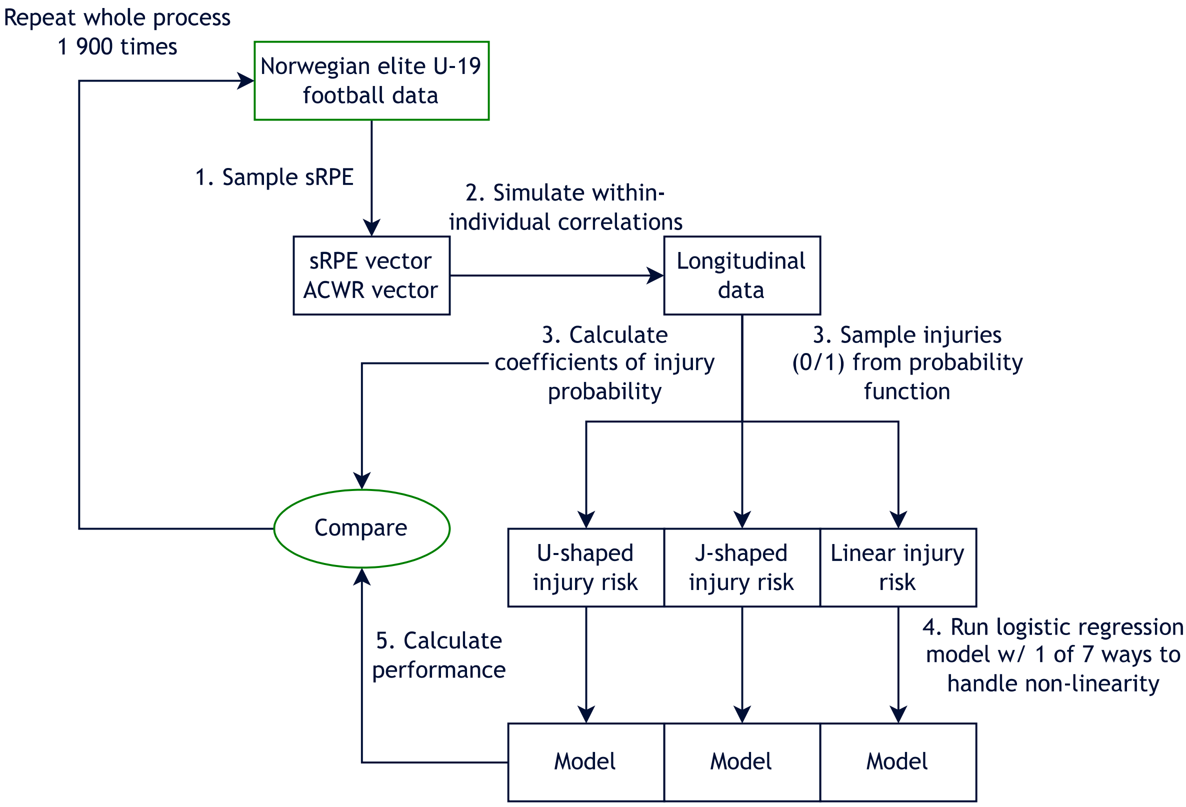

Figure 13. The five steps of the simulation workflow in *Paper II*: (1) sample training load values from the elite U-19 football data; (2) simulate correlations between training load values on the same individual; (3) calculate coefficients of injury probability according to three different training load/injury relationships; (4) fit one of seven different models with injury as the outcome and training load as the explanatory variable; (5) calculate performance measures and compare predicted probabilities with the simulated probabilities.

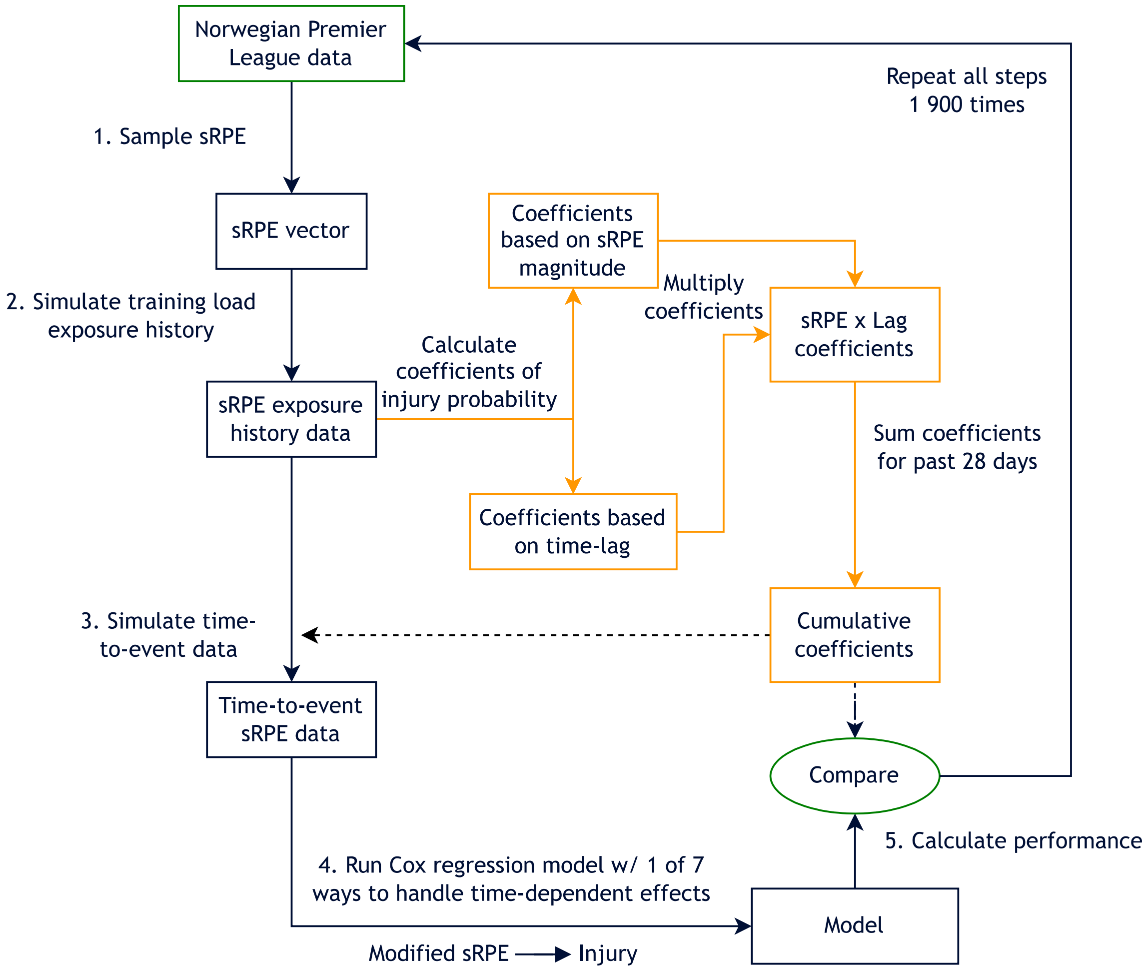

Figure 14. The five steps of the simulation workflow in *Paper III*: (1) extract training load exposure measured by session Rating of Perceived Exertion (sRPE) from the Norwegian Premier League dataset and (2) simulate training load exposure for 250 participants across 300 days; (3) calculate 1 of 7 scenarios of injury probabilities based on the cumulative training load observed the last 28 days, a combination of effect from both the magnitude of the training load (level of sRPE or %ΔsRPE) and the time since the training load occurred. Simulate injuries based on these probabilities to generate time-to-event data. (4) modify the absolute and relative training load exposures with seven different functions in Cox regression models; (5) calculate performance measures by comparing model with the simulated coefficients.

2.6.1 Preparing data

The Norwegian Premier League football data was used in *Paper I* and *Paper III* (4 725 sRPE and 2 292 total distance values of approximately Gaussian distributions, *Paper I* Figure SI). In *Paper I*, one dataset was formed with the sRPE and other variables, and one was formed with total distance and other variables. The correlations between all variables in the datasets were retained (*Paper I* Table 1). This was to ensure that methods of handling missing data that use other variables in the dataset as predictors for imputed values had realistic predictive ability, and potential important predictors could be identified. In *Paper III*, we sampled sRPE values with replacement to generate a faux study of 250 participants (10 football teams), followed for one full season (300 days). In addition to absolute training load, the relative training load from one day to the next was calculated with the symmetrized percentage change (%ΔsRPE, Curran-Everett & Williams, 2015), ranging from -100% to 100%.

In *Paper II*, the simulations were based on the Norwegian elite U-19 football dataset. Two datasets were used. One with the original 8 495 sRPE and 6 308 ACWR values, and one created by sampling sRPE and ACWR values with replacement to generate a scenario of 3 football teams (75 players) followed for a season (300 days; n training load values = 22 500). We simulated a longitudinal design with an autoregressive correlation structure. This structure imposes stronger correlations between observations closer in time than those more distant in time, which is considered realistic of longitudinal sports data on human participants (Wilkinson & Akenhead, 2013). In both datasets, the sRPE distribution was highly skewed to the right, as 0 was by far the most common value. The remaining distribution centered around an sRPE of 500, but many values were spread out in the 800 to 1500 range (*Paper II* Supplementary Figure S4). The ACWR distribution was approximately Gaussian (*Paper II* Figure S4).

2.6.2 Simulating a relationship between training load and injury

Simulated events (1/0) were added to the prepared datasets with a predefined, probabilistic relationship with training load. Only one event was simulated per individual. We used the term injury to describe the simulated events. However, the events can also be considered occurrences of pain or other health problems.

2.6.2.1 Logistic regression (*Paper I–II*)

A linear relationship between sRPE and probability of injury was simulated in *Paper I–II*, with the logit link function.

$$logistic(x) = \frac{1}{1 + \exp(-x)}$$

$$Prob(Injury) = logistic(\beta_0 + \beta_1 x)$$

Where $\beta_0 + \beta_1 x$ was defined as

Paper I $-2 + 0.003 * sRPE$

$-2 + 0.0003 * Total\ Distance$

Paper II $-0.5 + \ 0.001 * sRPE$

In addition, to test the ability of methods for detecting non-linear relationships, two additional relationships were simulated in *Paper II*.

U shape. A symmetrical parabola with sRPE (Gamble, 2013).

$$Prob\{Y = 1|sRPE\} = \ logistic(-1 + \ 0.0000002 * (sRPE - 1500)^2)$$

J shape. An asymmetrical parabola with ACWR (Blanch & Gabbett, 2016; Carey et al., 2018).

$$Prob\{Y = 1|ACWR\} = logistic(\begin{cases} -3.4 + 2 * (1 - ACWR)^2, & ACWR < 1 \\ -3.4 + (1 - ACWR)^2, & 1 \leq ACWR < 1.7 \\ 1.5 * ACWR - 5.4, & ACWR \geq 1.7 \end{cases})$$

Altogether, two relationships were simulated in *Paper I*, and three relationships in *Paper II*. After simulating relationships in *Paper II*, noise was added to training load values with the default jitter value in the statistical program to mimic measurement error.

2.6.2.2 Cox regression (*Paper III*)

In *Paper III* we considered a time-to-event approach and simulated survival histories (time to injury) using the Cox regression model:

$$h(t) = h_0(t) * \exp(\beta x)$$

Where h_0 is the baseline hazard, and $h(t)$ is the hazard at timepoint t. The coefficient β was the result of a bidimensional function s on both the magnitude of the training load x, and the distance in time, the time lag l, from the timepoint t.

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

Training load x was measured with the sRPE. 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. The current day, Day 0, was $l = 0$. The max lag was $L = 27$, in other words, 28 days (4 weeks).

For every day between $l = 0$ and $l = 27$ we ran a function $f(x)$ on the magnitude of training load, and function $w(l)$ on the time since the current day. We simulated s to be a cumulative sum of these 28 results, moving iteratively from one day to the next.

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

The relationship between the magnitude of training load and probability of injury $f(x)$ was simulated with two different functions (*Paper III* Figure S2A, S2C).

J shape. For absolute training load. $f(x) = \begin{cases} ((600 - x)/200)\hat{}1.5/10, & x < 600 \\ ((x - 600)/200)\hat{}3/30), & x \geq 600 \end{cases}$

Linear shape. For relative training load. $f(\%\Delta x) = 0.009 * \%\Delta x$

The relationship between the time since current day and probability of injury $w(l)$ was simulated with four different functions, corresponding to various hypothetical scenarios (*Paper III* Figure S3A–D).

Constant. Across 4 weeks, the effect of training load was constant each day.

$$w(l) = 0.8$$

Decay. Across 4 weeks, the effect of training load gradually decayed for each day.

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

Exponential decay. The effect of training load dropped exponentially during the past 4 weeks.

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

Direct, then inverse. Training load values on the current week (acute) increased risk of injury, whilst the training load values three weeks before the current week (chronic) decreased risk of injury (Blanch & Gabbett, 2016).

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

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 (Gabbett, 2016; C. Wang, T. Stokes, R. Steele, et al., 2021; C. Wang, T. Stokes, J. T. Vargas, et al., 2021). In addition, for this time-lag scenario, and for this time-lag scenario only, we simulated a linear relationship with the absolute training load (*Paper III* Figure S2B):

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

All in all, in *Paper III,* seven different relationships between training load and injury risk were simulated, four with absolute training load and three with relative training load (*Paper III* Figure 1–2). A censoring timepoint was drawn at random from a uniform distribution ranging from 0 to 600 days per individual. The mean number of simulated injuries for 25 participants (a football team) across 100 simulations for each of the seven scenarios, was 18.7 per season; reasonably realistic of a study with small-to-moderate effect between training load and a specific injury type (i.e. a study on hamstring injury).

2.6.3 Simulating missing data

In *Paper I*, we also simulated missing data. From the sRPE dataset and total distance dataset – now with simulated relationships with injury – eleven datasets were created with amounts of missing sampled under the assumption of Missing Completely at Random (MCAR): 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%, to cover a range of percentages of missing data (Vink, 2016).

We also created three datasets under the assumption of Missing at Random (MAR).
The probability of missing in sRPE or total distance under MAR was based on the following functions:

Light MAR (~25% missing)

$$Prob(Missing) = logistic(-2 + 0.03 * Age + \; 0.02 * Sex + 0.3 * Recovery\ day)$$

Medium MAR (~50% missing)

$$Prob(Missing) = logistic(-2 + 0.08 * Age + \; 0.04 * Sex + 0.8 * Recovery\ day)$$

Strong MAR (~80% missing)

$$Prob(Missing) =$$

$$logistic(-2 + 0.13 * Age + \; 0.1 * Sex + 1.8 * Recovery\ day + 1.8 * Match)$$

Simulated age values were drawn at random (uniformly) from 18 to 30 years and added to the dataset. Simulated sex values were randomly sampled from female 0 and male 1, with probability 50% for each sex. A recovery day (0/1) was the day after a match or two days after a match (M+1 or M+2), and match was also coded 0/1 (available in the observed data).

In summary, 11 MCAR datasets and 3 MAR datasets were generated for the sRPE and total distance datasets (24 datasets).

2.6.4 Choosing statistical methods for comparison

Several methods were compared in their ability to handle missing data (*Paper I*), handle non-linearity (*Paper II*), and handle the cumulative protracted time-lagged effect of long-term training load (*Paper III*). We chose statistical methods of comparison on the following bases:

- Methods frequently used in the training load and injury risk field. In *Paper I*, we used our systematic review of the field of training load and injury risk for an estimate of the most popular methods.
- Methods recommended in the training load and injury risk field.
- Methods we considered having potential in the corresponding scenario.

Methods for handling missing data

Strategies of imputing a derived variable

sRPE is a derived variable, the product of RPE and activity duration in minutes. We analyzed how sRPE should be imputed. After simulating a relationship between sRPE and injury, the sRPE variable was removed from the dataset, and 25% of RPE and duration observations were imputed completely at random, using multiple imputation with predicted mean matching, under four different strategies.

Impute, then transform. Impute duration and RPE without sRPE in the dataset, and calculate sRPE after imputation (Von Hippel, 2009). With this method, the product, sRPE, is not available to inform the imputation model. However, it may reduce collinearity issues.

Transform, then impute. Calculate sRPE, and impute duration, RPE and sRPE as regular variables (Von Hippel, 2009; White et al., 2011). Here, sRPE is present to inform the imputation model.

Passive imputation. Calculate sRPE and impute, but add the relationship between RPE, duration and sRPE in the imputation model, thereby transforming on-the-fly within the imputation algorithm (Van Buuren, 2018). This may be an improvement over using them merely as explanatory variables.

Impute product without factors.[2] Calculate sRPE, remove RPE and duration from the dataset, then impute. Under this scenario, the factors, RPE and duration, are not available to inform the imputation model, but it may reduce collinearity issues. This may be reasonable for studies which only have access to the product, sRPE.

The amount of bias introduced from the strategies were compared. The most accurate method determined in these simulations were used in the main comparison.

Main comparison

The missing observations in the 14 sRPE and 14 total distance datasets were imputed or deleted with five different methods.

[2]This approach was used to impute sRPE in datasets used in this dissertation before *Paper I* was conducted.

- Complete case analysis
- Mean imputation by the mean per player
- Mean imputation by the mean per week
- Regression imputation
- Multiple imputation with predicted mean matching (PMM)

Regression imputation and PMM uses the other variables in the dataset to predict training load observations in place of missing values. In both cases, we used a linear model, and the variables used as predictors are listed in *Paper I* Table 1. PMM draws at random from a pool of so-called donor observations, that are predicted to be most similar to the missing value. We used 5 donors, and 5 datasets to be imputed in the multiple imputation framework (Van Buuren, 2018). In addition, we compared PMM run with single and multiple imputation, to have an idea of how much of its performance stems from the multiple imputation framework.

Logistic regression models were run with training load, either sRPE or total distance, as the exposure, and the simulated injuries as the outcome variable, on each of the imputed 14 sRPE and 14 total distance datasets.

Imputation with extra variables available

To see how imputation models used in regression imputation and multiple imputation perform with and without certain variables, the simulation was repeated under different scenarios.

1. Only total distance was imputed, and no extra variables were in the imputation model.
2. Only total distance was imputed, and the player's playing position was among the variables in the imputation model.
3. Only total distance was imputed, and both the player's playing position and the sRPE was among the variables in the imputation model.
4. All GPS variables were imputed, and no extra variables were in the imputation model.
5. All GPS variables were imputed, and the player's playing position was among the variables in the imputation model.
6. All GPS variables were imputed, and both the player's playing position and the sRPE was among the variables in the imputation model.

Methods

Methods for handling non-linearity

A mixed effects logistic regression model was used to estimate the relationship between training load and predefined injury probability. Seven different model specifications of training load were compared in their ability to detect this predefined relationship.

Linear Model. A standard logistic regression was run to determine the magnitude of error that can be the result of assuming linearity when the relationship is non-linear. In a logistic regression with x_1 representing the training load variable, the formula was as follows:

$$Prob\{Y = 1|X\} = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \gamma\right)}{1 + \exp\left(\beta_0 + \beta_1 x_1 + \gamma\right)} = logistic(\beta_0 + \beta_1 x_1 + \gamma)$$

Where γ was the random effect term.

Categorization. We attempted to reproduce the results of categorization in Carey et al. (2018) in an environment of highly skewed sRPE values. To determine whether results can differ depending on how the data are categorized, we used two different approaches to choose the cut-off values for the categories. In one, the data-driven approach, training load was delineated by quartiles (like in Cross et al., 2016; Malone et al., 2019; Stares et al., 2018). In the other, the subjective approach, cut-offs were decided based on the range of the data.

sRPE was parted in four categories.

- <= 499
- 500–1 499
- 1 500–2 499
- >= 2 500

ACWR was parted in three categories, same as in Carey et al. (2018).

- < 1
- 1–1.74
- >= 1.75

Quadratic model. To assess whether quadratic regression is sufficiently accurate if the relationship is U- or J-shaped, a quadratic model was among the compared methods. In a quadratic model, the explanatory variable is modeled with a polynomial to the second power.

$$Prob\{Y = 1|X\} = logistic(\beta_0 + \beta_1 x_1 + \beta_2 {x_1}^2 + \gamma)$$

Fractional polynomials. Fractional polynomials add either a single polynomial term (FP1) or two polynomial terms (FP2) to the pth power to the regression model (Royston & Altman, 1994). FP2 was used in all models in *Paper II*, as it is the most optimal choice in most cases (Binder et al., 2013). The logistic regression model with FP2 was as follows:

$$Prob\{Y = 1|X\} = \ logistic(\beta_0 + \ \beta_1 x_1 + \ \beta_2 {x_1}^{p1} + \beta_3 {x_1}^{p2} + \gamma)$$

Where $p1$ and $p2$ are exponents selected from {-2, -1, -0.5, 0, 0.5, 1, 2, 3} by a form of backward elimination (Ambler & Benner, 2015). If $p1$ or $p2$ = 0, x^p is replaced with $\ln(x)$.

Restricted cubic splines. In restricted cubic splines (RCS), the X-axis is divided into intervals by a number of endpoints (knots). At these knots, different cubic polynomials are joined and forced to have a consistent function, slope and acceleration (second derivative) until the next knot. At the knot, the rate change of acceleration (third derivative) may change. For three knots a, b and c, our logistic regression formula was:

$$Prob\{Y = 1|X\} =$$

$$logistic[\beta_0 + \ \beta_1 x_1 + \beta_2 {x_1}^2 + \beta_3 {x_1}^3 + \beta_4 (x_1 - a)^3 + \beta_5 (x_1 - b)^3 + \beta_6 (x_1 - c)^3 + \ \ \gamma]$$

In all simulation models, 3 knots were used (Harrell Jr, 2017). We compared two different ways of choosing knot location. In the first, the knot locations were chosen by the default approach in the statistical software (data-driven), and in the other, knot locations were cut-off subjectively at sRPE = 500, 1 500 and 2 500, and likewise at ACWR = 1, 1.75 and 2, to cover the range of the training load measures.

Methods for handling cumulative, protracted, time-lagged effects

A Cox regression model was used to estimate the relationship between training load and predefined injury probability. Seven different modifications or model specifications of training load were compared in their ability to detect the seven scenarios of cumulative, protracted time-lagged effects of training load on injury risk, four for the absolute training load x, and three for the relative training load $\%\Delta x$. To compare methods under the same conditions, absolute 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. Relative training load was modelled with a linear relationship with injury risk.

Some methods could not be used on incomplete time-windows. We therefore did not include the initial 27 days per individual when performing all methods, to improve comparability.

Absolute training load

RA. For training load denoted x, the rolling average RA is defined by:

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

Essentially, the RA is the mean of each time-lag window, where n is the size of the time-lag window (28 days). The calculation moved iteratively, one day at a time, in a sliding window. If k is the last value in the time-lag window for an individual, the first window was k = 28 (the first 4 weeks), the second window, $k = 29$, and so on, up until k = 300.

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

EWMA. The exponentially weighted moving average (EWMA) is calculated in the same way as the RA, but with a weighting term λ on the time since exposure by number of days n, up to a maximum of $n = 28$.

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

We used the same definition of λ as in Williams, West, et al. (2017) and Moussa et al. (2019).

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

REDI. The Robust Exponential Decreasing Index (REDI), developed by Moussa et al. (2019), multiplies a vector of coefficients (weights) with the vector of training load values, for the lag interval $l = 0, \ldots, L$ where $l = 0$ is the current day, and L is the maximum lag 27. These weighted training load values are subsequently summed.

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

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

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

We assumed the observations had been imputed before-hand, and hence, $\alpha_l^\lambda = \exp(-\lambda * l)$ for every value. The weighted training load values were then divided by the sum of the weight coefficients:

$$REDI = \frac{1}{\text{Weighted x}} * \sum_{l=0}^{\text{L}} \alpha_l^\lambda$$

We chose lambda = 0.1, same as in Moussa et al. (2019), since it was the highest lambda value where training load on the 27th lag day still contributed to the cumulative effect.

DLNM. We previously described how the β-coefficient for training load can be a result of the s function, $s(x_t, \ldots, x_{t-L})$. By defining s as the product of the variable function $f(x)$ and the lag function $w(x, l)$, we can consider s as the result of bi-dimensional exposure-lag-response function $f(x) \cdot w(x, l)$ in a distributed lag non-linear model (Gasparrini, 2014):

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

We modelled $f(x)$ with a quadratic term in the Cox regression model, except for the "Direct, then inverse" time-lag scenario, where we used a linear term instead (same as for the other methods). We modelled the lag-response function $w(x, l)$ with restricted cubic splines using 3 knots under all scenarios.

Relative training load

Week-to-week %-change The percentage difference in weekly sRPE sums between the current week and the previous week (Ramskov et al., 2021; Ryan et al., 2021). We used the symmetrized percentage change (Curran-Everett & Williams, 2015):

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

Where W is the sum of daily sRPE across 7 days, and k is the current week. The week-to-week percentage change calculation moved iteratively from one week to the next.

7:28 coupled ACWR. The traditional Acute: Chronic Workload Ratio (ACWR) was the most common form of calculation in a systematic review of ACWR in male football research (A. Wang et al., 2021). This was the sum of training load on the current week (Day 6 up to Day 0)—the acute training load—divided by the rolling average of the current week and the previous three weeks (Day 27 up to Day 0).

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

Where k is the current week. The ACWR calculation was moved iteratively from one day to the next (Carey et al., 2017).

DLNM. In addition to the absolute training load, DLNM was also compared on the relative training load. The exposure-response function $f(\%\Delta x)$ was modelled linearly, same 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.

2.6.5 Assessing performance

The final step in all simulation studies was to assess the performance of compared methods.

In *Paper I*, we assessed the validity of the imputation by visually comparing the imputed with the observed data for 50% missing under MCAR and ≈80% missing under MAR. Ideally, the imputed data should follow the same distribution as the observed data.

After running a logistic regression model on imputed datasets in *Paper I*, the percentage bias was calculated to determine accuracy of detecting a relationship between training load and injury risk. It was based on the absolute percentage difference between the estimated coefficients and the true coefficients used to simulate injuries. The upper limit for acceptable performance was $|5\%|$ (Demirtas et al., 2008). The percent bias per method was visualized for each scenario of missing.

From the absolute bias, we derived the Root-Mean-Squared Error (RMSE). The RMSE is a combined metric of accuracy and precision, and was the main numeric performance measure in *Paper II* and *Paper III*, where it was calculated between the simulated (true) probability of injury and the probability predicted by the compared models.

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

RMSE is used to rank methods—the lower the RMSE, the better the method. However, the scale of the RMSE depends on the analysis in question, and the values cannot be interpreted in isolation (Morris et al., 2019). In both papers, RMSE was arranged from highest to lowest in dot plots.

The ACWR and week-to-week percentage change methods for assessing relative training load in in *Paper III* distorted the coefficients to a different scale, and the difference between estimated and true risk could not be calculated. We therefore also included the RMSE calculated on the residuals (internal RMSE), and Akaike's Information Criterion (AIC) as alternative measures of model fit in *Paper III*. The internal RMSE and AIC were also visualized in dot plots.

To compare the uncertainty of measures, coverage of 95% confidence intervals (CIs) and average width (AW) of 95% CIs were assessed in *Paper I* and *Paper III*.

In *Paper II*, predictive ability and model fit as described by Brier scores, C-statistics, and 95% prediction intervals (PIs) were also considered.

The number of simulations needed for an accurate estimation of coverage was estimated based on a Monte Carlo Standard Error of 0.5 (Morris et al., 2019):

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

The simulations in *Paper I–III* were repeated 1 900 times for all methods and relationship scenarios. The mean of each performance measure was calculated across these 1 900 simulations.

Visualizations were made in *Paper II* and *Paper III* with the simulated risk of injury for each level of sRPE, compared with risk of injury estimated by the different methods. Only 1 of the 1 900 simulations were chosen at random to be displayed in these figures.

2.7 Statistical analysis 3: Observed sports data

We assessed whether there were any signs of non-linearity, cumulative protracted time-lagged effects, and/or interactions in the relationship between training load and injury risk, in different sport populations. All Norwegian populations were assessed in *Paper II*, the Norwegian elite youth handball was assessed again in *Paper III*, and the Norwegian elite U-19 football (again) and Qatar Stars League football were assessed in *Paper IV* (Table 1).

Table 1. Overview of models exploring complex effects in the relationship between training load and injury risk.

Study	TL measures	Model	Non-linearity	Time-lagged effects	Interaction
Paper II					
Norwegian Premier League	sRPE, daily ACWR, micro-cycle ACWR	Logit	X		
Norwegian elite U-19	sRPE, daily ACWR, micro-cycle ACWR	Logit	X		
Norwegian elite youth handball	sRPE, daily ACWR, micro-cycle ACWR	Logit	X		
Paper III					
Norwegian elite youth handball	sRPE	Cox	X	X	
Paper IV					
Qatar Stars League	Minutes in activity	Logit	X	X	X
Norwegian elite U-19	sRPE	Logit	X	X	X

Abbrevations: ACWR = Acute:Chronic Workload Ratio; sRPE = session Rating of Perceived Exertion; TL = Training load

2.7.1 Training load measure modification

In *Paper II*, we investigated whether there were any tendencies of non-linearity between training load and probability of injury in all three Norwegian sports populations. Since the shape of the relationship between training load and injury is potentially different for absolute versus relative training load, we calculated the relative training load in addition to the absolute training load. The relative training load was calculated with the Acute:Chronic Workload Ratio (ACWR); both daily and micro-cycle ACWR.

Daily uncoupled ACWR 7:21 The mean sRPE of 7 days (acute load) divided by the exponentially-weighted-moving average (EWMA) of the previous 21 days (chronic load). The acute load was not included in the 21 days of the denominator (uncoupled ACWR, C. Wang et al., 2020). The calculation was performed on a sliding window moving iteratively one day at a time from and including the 28th day (Carey et al., 2017). The last day in the acute load was considered the current day (Day 0).

Micro-cycle ACWR 1:3 The mean sRPE per micro-cycle divided by the EWMA of the previous 3 micro-cycles, uncoupled. A micro-cycle was all recovery days after the previous match, and the training days before the next match (Figure 8). The calculation was performed in the same manner as daily ACWR, though on a sliding window moving one micro-cycle at a time from and including the 4th micro-cycle.

2.7.2 Statistical analysis

To model the association between training load and injury risk, we used mixed-effects logistic regression (Nielsen, Shrier, et al., 2020) in *Paper II* and *Paper IV*, and in *Paper III*, Cox regression with frailty (Nielsen et al., 2019; Ullah et al., 2014). The random intercept terms and the frailty term were modelled per handball or football player in the data, to account for within-player dependencies. Injury (yes/no) was the outcome in all models.[3] Days where players were not at risk (sRPE = 0) were removed before analysis, as recommended in Mustapich and Koehle (2021). See Table 1 for an overview of the models.

Training load was the independent variable. Absolute training load measured by daily sRPE was used in all cases, expect the Qatar Stars League model, which only had the daily activity duration in minutes (*Paper IV*). Relative training load, as calculated by the daily ACWR and micro-cycle ACWR, was additionally used in *Paper II*. In all cases, training load was modelled with restricted cubic splines using three knots. In *Paper II* and *Paper III*, the knots were located at quartiles of training load measures, in *Paper IV*, they were subjectively placed based on the range of the sRPE values. The models were repeated without splines in *Paper II* to determine the relationship we would have discovered if linearity was assumed.

The potential effect of past (chronic) training load was analyzed differently in the three papers. *Paper II* considered the effect of training load sustained five days ago (Lag day 4, or day -4) on the occurrence of injury during the next four days (Day -3 to Day 0), where the training observation day (Day -4) was not included (*Paper II* Figure 1). For micro-cycle ACWR, we estimated the association of relative training load in the previous micro-cycle with the risk of injury occurring during the next micro-cycle excluding Day 0 (*Paper II* Figure 1).

To explore the potential for cumulative, protracted, time-lagged effects of past training load in *Paper III* and *Paper IV*, the last four weeks (28 days) of sRPE was modelled with the distributed lag non-linear model described in 2.6.4 (Gasparrini, 2014). The lag-response function $w(x, l)$ was modelled using a restricted cubic splines with three knots in the Norwegian elite youth handball model (*Paper III*) and Norwegian elite U-19 football model (*Paper IV*), and with four knots in the Qatar Stars League model (*Paper IV*), as it had a high sample size. Although risk-free days (sRPE

[3] In *Paper III*, we erroneously wrote that we studied all health problems in the Norwegian elite youth handball model, although we only studied injuries.

= 0) were not included in the model, they were included in the DLNM calculation of past training load.

Day 0 was not included in the last four weeks of sRPE in *Paper IV*. Instead, it was included as a separate variable in the model and considered the acute load. The past training load modelled with DLNM was considered the chronic load. An interaction was added between the acute load and the chronic load. In addition to the DLNM, the EWMA method was run on the chronic load to see if a simpler approach was appropriate.

Age and sex were adjusted for in the Norwegian elite U-19 football (*Paper II*) and Norwegian elite youth handball models (*Paper II-III*), and age was adjusted for in the Norwegian Premier League model (*Paper II*).

The main result in all three papers was a visualization of the model predictions to determine the shape of the relationship between training load and injury risk, in line with recommendations in Shrier et al. (2021). Exploration of the effects of chronic training load were also visualized in *Paper III* (Norwegian elite youth handball) and *Paper IV* (Qatar Stars League model only).

2.8 Data tools

Data preparation, statistical analyses and simulations were performed using R (R Core Team, 2021). Aside from packages in base R and in the Tidyverse family (Wickham, 2019; Wickham et al., 2019)—used for handling, reading and plotting of data—other packages were used for specific purposes (Table 2). A GitHub repository with R code and study protocols was made publicly available for each study: *Paper I* (Bache-Mathiesen, 2021a), *Paper II* (Bache-Mathiesen, 2021b), *Paper III* (Bache-Mathiesen, 2022b), *Paper IV* (Bache-Mathiesen, 2022a). Infographics and flowcharts were made in diagrams.net v.20.2.7 (Alder, 2018).

Table 2. R packages used in this dissertation.

Package	Purpose	Reference
chron	Manipulating time data	James (2020)
clubSandwich	Cluster-robust confidence intervals	Pustejovsky (2021)
directlabels	Labeling line graph lines	Hocking (2021)
DLNM	Distributed lag non-linear models	Gasparrini (2011)
doParallel	Running multiple cores simultaneously	Weston and Microsoft (2022a)
egg	Labeling plots in a panel	Auguié (2019)
foreach	For-loops run in doParallel	Weston and Microsoft (2022b)
ggeffects	Model predictions	Lüdecke (2018)
lme4	Generalized mixed effects models	Bates et al. (2015)
merTools	Prediction intervals for mixed models	Knowles and Frederick (2019)
mfp	Fractional polynomials	Ambler and Benner (2015)
mice	Multiple imputation	Buuren (2011)
PermAlgo	Simulating time-to-event data	Sylvestre and Abrahamowicz (2008)
rms	Restricted cubic splines	Harrell Jr (2019)
SimCorMultRes	Simulating longitudinal data	Touloumis (2016)
sjPlot	Plotting splines predictions	Lüdecke (2022)
slider	Functions on sliding windows	Vaughan (2021)
tsModel	Structuring time series data	Peng and McDermott (2022)
TTR	Exponentially weighted moving averages	Ulrich (2020)
visdat	Visualizing missing data	Tierney (2017)
zoo	Rolling averages	Zeileis and Grothendieck (2005)

3 Results

3.1 Review results

3.1.1 Current practices of handling missing data in training load

In our review of 108 studies, 34% reported whether the training load variable had any missing observations. This varied between 30%–50% the last five years (*Paper I* Figure 2). Fewer studies (23%) reported how they handled the missing data (Table 3), and fewer still (17%) reported the amount of missing data. The mean percentage missing was 7% (SD = 6%). Mean imputation (n = 11) and complete case analysis (n = 8) were the most frequently used methods for handling missing data (Table 3).

Table 3. The methods used to handle missing data in training load in the field of training load and injury risk (n = 36[1]).

Missing Data Method	N studies	% of studies
Unclear[2]	12	33%
Mean Imputation	11	31%
Complete Case Analysis	8	22%
Median Imputation	2	6%
Multiple Imputation	2	6%
Regression Imputation	1	3%

[1] Although 37 (34%) of 108 studies reported whether they had missing data in the training load variable, one of the 37 studies had no missing data, and therefore removed from this analysis.
[2] Cases were defined as "unclear" if authors reported having missing data, but the method used to handle them were unclear.

3.1.2 Sample sizes in training load and injury risk studies

Most studies reviewed in *Paper I* were conducted across 1 season/year/school year of the target population (52% of 108 studies). See *Paper I* Table 2 for study characteristics. The number of analyzed injuries followed a right-skewed distribution (Figure 15). The median number was 85, with a 25[th] and 75[th] percentile at 36 and 159 injuries, respectively.

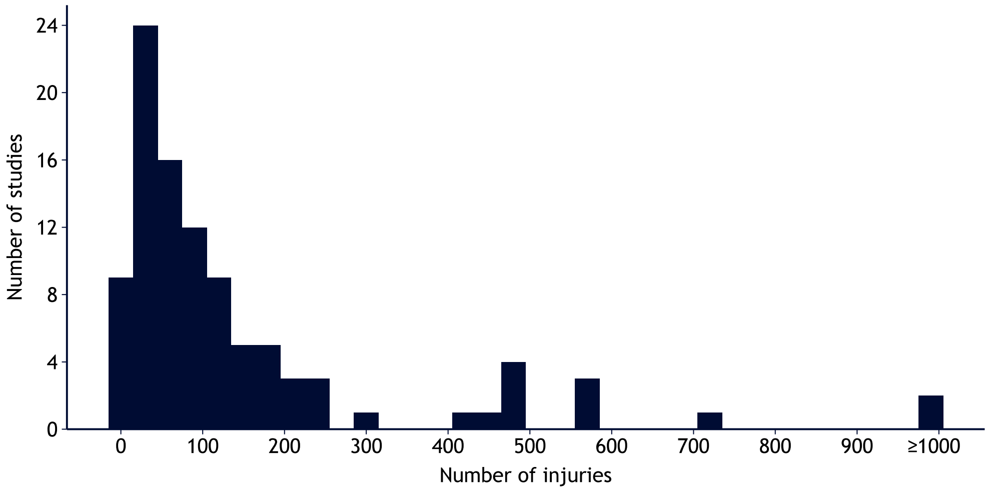

Figure 15. The number of injuries analyzed in training load and injury studies formed a right-skewed distribution. Descriptive statistics: Median = 85, 25th percentile = 36, 75th percentile = 159, interquartile range = 123. Based on 99 of 108 studies that reported the number of registered injuries.

3.2 Simulation results

3.2.1 Handling missing data in session Rating of Perceived Exertion

In the pilot simulation, "Impute, then transform" had the lowest bias (1.4%) and "Impute product without factors" the second-lowest (1.8%) of the four approaches of imputing the compound variable sRPE (Table 4).

In the main simulation, multiple imputation using Predicted Mean Matching (PMM) had the lowest average bias across all proportions of missing data under MCAR (6% vs. >= 10% [all other methods]) and a sufficient coverage of 95% (Table 5). It was the only method with acceptable bias (< |5%|) up to 50% missing (Figure 16A).

Table 4. The mean percent bias, root-mean-squared-error, and average width of 95% confidence intervals across 1 900 simulations, for four approaches of imputing sRPE.

Imputation Method	% Bias	RMSE	AW
Impute, then transform	1.4%	0.0000124	0.000745
Transform, then impute	2.6%	0.0001100	0.000943
Passive imputation	2.5%	0.0000894	0.000713
Impute product without factors	1.8%	0.0000599	0.000895

Abbreviations: AW, Average width; sRPE, session Rating of Perceived Exertion; RMSE, Root-Mean-Squared-Error

Table 5. The mean performance for five methods of imputing/deleting missing data in session Rating of Perceived Exertion (sRPE) and total running distance. Calculated across 11 scenarios of MCAR (n = 20 900) and 3 scenarios of MAR (n = 5 700). Compared to performance without missing data (None).

Missing	Missing variables[1]		PB[2]	RMSE[2]	Coverage[3]	AW
None			1.4%	0.000042	100%	0.000624
MCAR	sRPE	Complete Case Analysis	10%	0.000319	95%	0.001910
		Mean per player	11.4%	0.000357	80%	0.000938
		Mean per week	10.4%	0.000338	65%	0.000922
		MI – PMM	5.8%	0.000191	95%	0.001400
		Regression Imputation	33.7%	0.001040	30%	0.000828
MAR		Complete Case Analysis	7.7%	0.000280	100%	0.001699
		Mean per player	7.9%	0.000254	100%	0.000991
		Mean per week	8.8%	0.000275	100%	0.000968
		MI – PMM	3.8%	0.000144	100%	0.001112
		Regression Imputation	38.8%	0.001175	33%	0.000833
None			4%	0.000012	100%	0.000081
MCAR	All GPS variables	Complete Case Analysis	8.2%	0.000025	100%	0.000126
		Mean per player	8.6%	0.000026	90%	0.000103
		Mean per week	13.1%	0.000039	90%	0.000099
		MI – PMM	10.7%	0.000032	87%	0.000082
		Regression Imputation	24.8%	0.000075	40%	0.000091
	Total distance only	Complete Case Analysis	8.9%	0.000027	93%	0.000122
		Mean per player	9.6%	0.000029	90%	0.000102
		Mean per week	18.2%	0.000055	70%	0.000098
		MI – PMM	3.3%	0.000010	100%	0.000081
		Regression Imputation	6.1%	0.000018	100%	0.000078
MAR	All GPS variables	Complete Case Analysis	20.3%	0.000061	78%	0.000156
		Mean per player	11.3%	0.000034	78%	0.000123
		Mean per week	11.7%	0.000035	89%	0.000120
		MI – PMM	9.5%	0.000029	78%	0.000198
		Regression Imputation	60.5%	0.000181	67%	0.000124
	Total distance only	Complete Case Analysis	14.1%	0.000042	100%	0.000143
		Mean per player	9.9%	0.000030	100%	0.000118
		Mean per week	11.4%	0.000034	89%	0.000114
		MI – PMM	6.7%	0.000020	100%	0.000098
		Regression Imputation	11.7%	0.000035	67%	0.000084

Abbreviations: AW, Average Width of 95% confidence intervals; GPS, Global Positioning System; MAR, Missing at Random; MCAR, Missing Completely at Random; PB, Absolute Percent Bias; sRPE, Session Rating of Perceived Exertion; RMSE, Root-Mean-Squared-Error

[1]All GPS variables = All GPS-variables have missing data, Total distance only = Only total distance has missing data

[2]Monte Carlo standard error < 0.0001

[3]Monte Carlo standard error = 0.5

Complete case analysis was within acceptable bias up to 20% missing, and the other methods had acceptable bias up to 10% missing (Figure 16A). PMM was the only method within acceptable bias under both the light (~25% missing) and medium (~50% missing) MAR scenarios (Figure 16B), at low cost to certainty (100% coverage, Table 5). Regression imputation was not within acceptable bias under any MAR scenarios (Figure 16B), while the other methods were within acceptable limits under light MAR (~25% missing).

PMM run in a single or a multiple imputation framework had varying results: For some levels of missing data, single imputation had the lowest bias, however for other levels of missing data, multiple imputation had the lowest bias (*Paper I* Table S1).

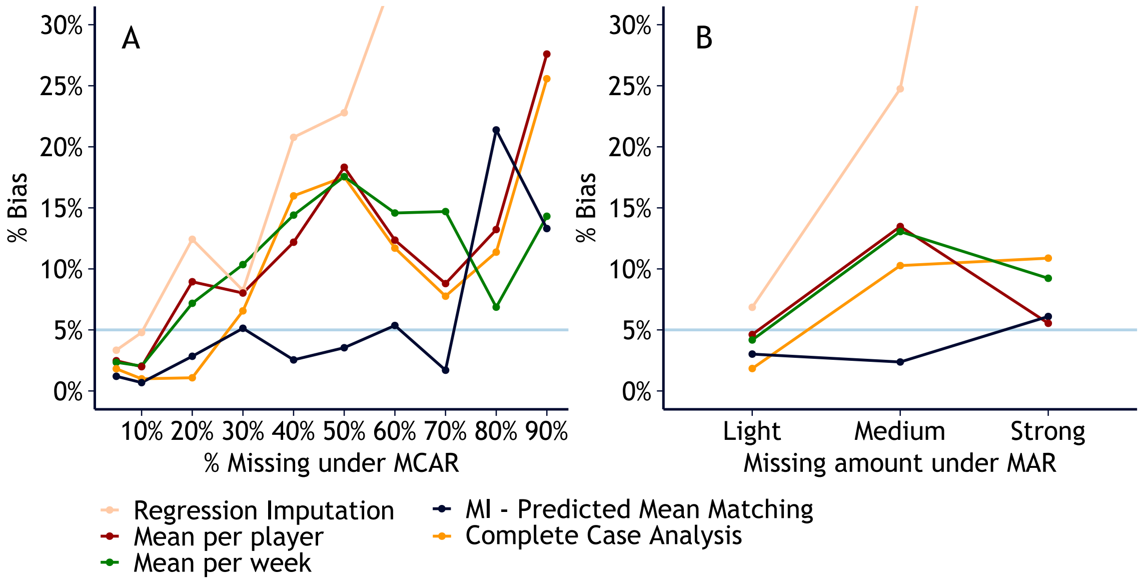

Figure 16. Accuracy of a logistic regression model after imputing or deleting missing observations in the session rating of perceived exertion (sRPE) with five different methods. Accuracy is measured by the mean absolute percent bias (% Bias) across 1 900 simulations. The X-axis displays varying levels of missing data under the assumption of (A) Missing Completely at Random (MCAR), and (B) Missing at Random (MAR). Under MAR, the probability of missing is dependent on other variables: Light (≈25% missing); Medium (≈50% missing); Strong (≈80% missing). The most accurate methods are closest to 0, and the light blue line indicates the maximum range of acceptable bias (0% to 5%). Methods off the chart had > 30% bias. Without missing data, the logistic regression had an inherent bias of 1.4%. Monte Carlo standard error < 0.00001.

3.2.2 Handling missing data in GPS-measures

Under MCAR, PMM was the only method with acceptable bias up to 90% missing in the GPS-measure of total distance run (3.3% mean bias across all proportions of missing, 100% coverage, Table 5). Complete case analysis was also acceptable up to 50% missing data (Figure 17A).

When all GPS variables were missing simultaneously, PMM was only within acceptable bias at 10% missing or less (*Paper I* Figure 6A). Here, complete case analysis and mean imputation by the mean per player was acceptable up to 20%, and mean imputation by the mean per week up to 30% missing data.

Mean imputation by the mean per player was also within acceptable bias up to and including ~50% missing data, in 4 out of 6 scenarios of MAR (*Paper I* Figure 7). PMM was also within acceptable bias up to ~50% missing when the total distance variable was the only variable missing (Figure 17B), but only up to ~25% missing if all GPS-variables were missing simultaneously (*Paper I* Figure 7A).

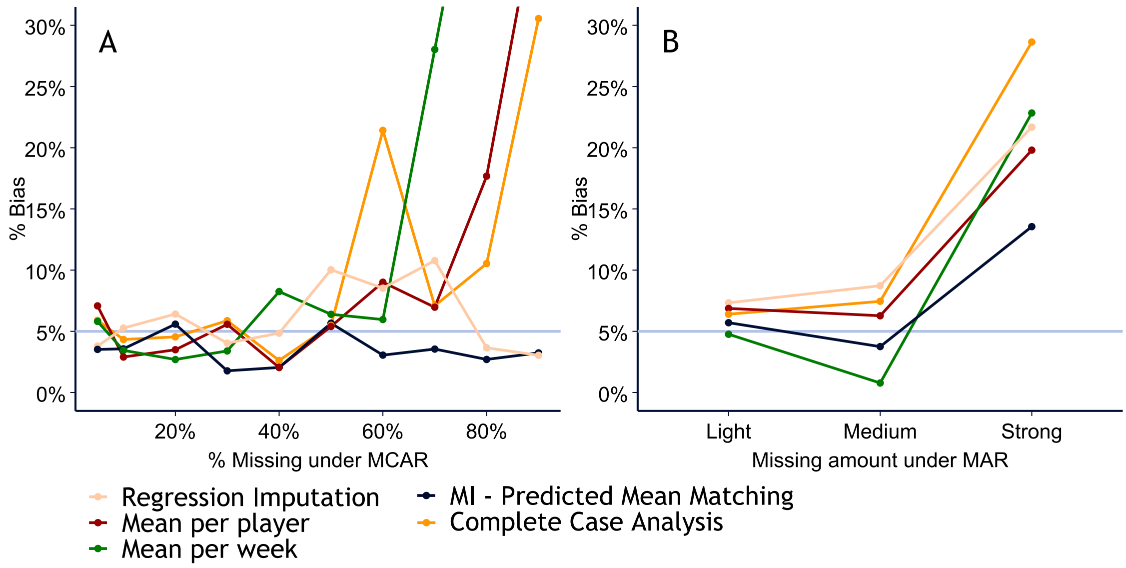

Figure 17. Accuracy of a logistic regression model after imputing or deleting missing observations in total distance with five different methods. Accuracy is measured by the mean percent bias (% Bias) across 1 900 simulations, under the assumption of (A) Missing Completely At Random (MCAR) and (B) Missing At Random (MAR). Under MAR, the probability of missing is dependent on other variables: Light (≈25% missing); Medium (≈50% missing); Strong (≈80% missing). The most accurate methods are closest to 0, and the light blue line indicates the maximum range of acceptable bias (0% to 5%). Methods off the chart had > 30% bias. Without missing data, the logistic regression had an inherent bias of 4%. Monte Carlo standard error < 0.00001. MI = Multiple Imputation.

Adding the player's playing position in the football team to the imputation model did not improve performance of PMM or regression imputation, but including the sRPE lowered bias of PMM and regression imputation under both MCAR and MAR (*Paper I* Figure 6–7).

3.2.3 Methods for addressing non-linearity

The quadratic model, fractional polynomials, and restricted cubic splines with subjectively placed knots were the only methods that accurately modeled the non-linear U-shaped relationship (Figure 18) and they had the lowest RMSE under this scenario (*Paper II* Figure 4A, Table 6). Here, restricted cubic splines with the data-driven approach had second-to highest RMSE.

For the J shaped relationship—the one based on ACWR—the quadratic model and fractional polynomials had the lowest RMSE (Table 6). Although both restricted cubic splines approaches had similar RMSE to the two categorization approaches (Table 6), the two categorization approaches had a coverage of prediction intervals at 79% and 89% respectively under n = 6 308, versus 94% and 90% for the two restricted cubic splines approaches, with similar coverage distributions under n = 22 500 (Table 6).

Categorization had poor coverage in general (Table 6), and categorizing by quartiles had particularly poor coverage for the linear shape (25% vs. > 99% for other methods, Table 6). Despite the poor coverage, under this scenario, categorizing by quartiles had a comparable brier score (0.24 vs ≈ 0.24 for other methods, n = 22 500) and C-statistic (0.59 vs. ~0.59 for other methods, n = 22 500). Similarly, the linear model could not form the U shape (Figure 18D) and had the highest RMSE for both non-linear shapes (Table 6), but it had a high C-statistic (> 0.8) for the U shape and moderate to poor C-statistic of the J shape (C-statistic = 0.77 for n = 6 308, C-statistic = 0.62 for n = 22 500).

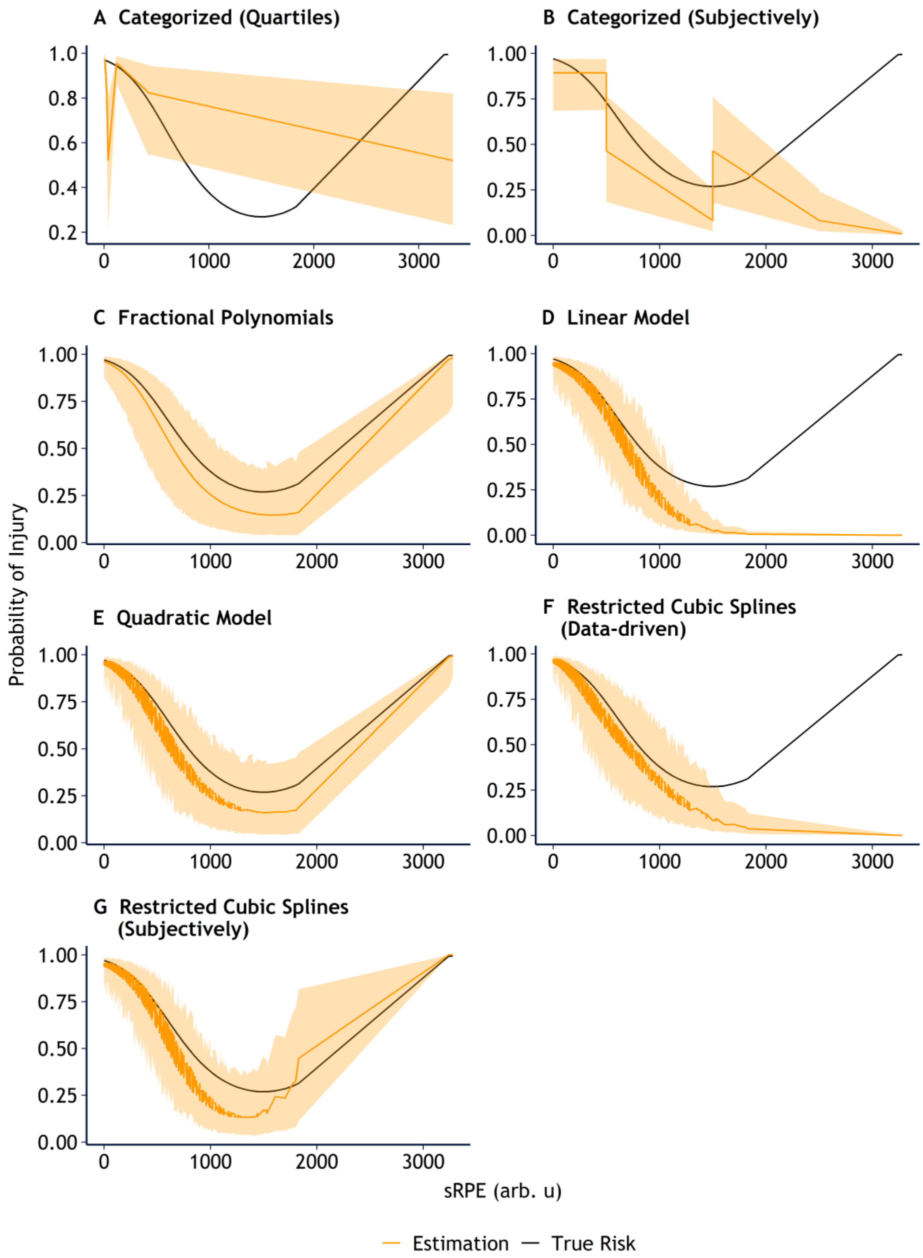

Figure 18. Probability of injury for each level of session Rating of Perceived Exertion (sRPE) as predicted by a logistic regression model run with seven different methods of model specification. The black line is the simulated, true, relationship between sRPE and injury—the yellow line represents the ability of the method to model this relationship. Ideally, the yellow line should follow the black line as closely as possible for the highest accuracy. The yellow area reflects the 95% prediction interval. The predictions are based on 8 494 sRPE values sampled from a highly skewed distribution in Norwegian elite U-19 football. The figure shows 1 random simulation of 1 900 performed. Arb. u = arbitrary units.

Table 6. The mean Root-Mean-Squared Error, Brier Score, C-Statistic and coverage of prediction intervals for methods modelling non-linear (U or J shape) and linear relationships between training load and injury, across 1900 simulations.

Relationship	N	Method	RMSE	Brier Score	C	Coverage
U shape	22 500	Linear Model	2.344	0.097	0.827	100.000%
		Categorized (Quartiles)	0.995	0.101	0.809	99.678%
		Categorized (Subjectively)	0.996	0.102	0.758	94.600%
		Quadratic Model	0.993	0.097	0.826	100.000%
		Fractional Polynomials	0.994	0.096	0.829	100.000%
		Restricted Cubic Splines (Data-Driven)	1.065	0.097	0.826	100.000%
		Restricted Cubic Splines (Subjectively)	0.981	0.097	0.827	100.000%
	8 494	Linear Model	2.935	0.093	0.851	98.048%
		Categorized (Quartiles)	0.958	0.096	0.838	98.769%
		Categorized (Subjectively)	0.965	0.098	0.809	84.600%
		Quadratic Model	0.956	0.092	0.850	98.937%
		Fractional Polynomials	0.956	0.092	0.852	98.942%
		Restricted Cubic Splines (Data-Driven)	1.079	0.092	0.849	98.686%
		Restricted Cubic Splines (Subjectively)	0.936	0.092	0.851	98.687%
J shape	22 500	Linear Model	1.044	0.063	0.618	77.694%
		Categorized (Quartiles)	0.993	0.064	0.689	88.652%
		Categorized (Subjectively)	0.993	0.063	0.690	96.404%
		Quadratic Model	0.984	0.061	0.732	99.997%
		Fractional Polynomials	0.986	0.061	0.740	100.000%
		Restricted Cubic Splines (Data-Driven)	0.992	0.061	0.735	99.999%
		Restricted Cubic Splines (Subjectively)	0.993	0.061	0.721	99.869%
	6 308	Linear Model	0.942	0.060	0.774	54.493%
		Categorized (Quartiles)	0.919	0.060	0.791	79.120%
		Categorized (Subjectively)	0.917	0.059	0.795	89.393%
		Quadratic Model	0.912	0.057	0.817	93.272%
		Fractional Polynomials	0.915	0.057	0.821	95.517%
		Restricted Cubic Splines (Data-Driven)	0.918	0.057	0.818	94.281%
		Restricted Cubic Splines (Subjectively)	0.919	0.057	0.812	89.959%
Linear	22 500	Linear Model	0.999	0.239	0.591	100.000%
		Categorized (Quartiles)	0.999	0.240	0.588	25.000%
		Categorized (Subjectively)	0.999	0.241	0.579	99.995%
		Quadratic Model	0.999	0.239	0.591	99.999%
		Fractional Polynomials	0.999	0.239	0.592	100.000%
		Restricted Cubic Splines (Data-Driven)	0.999	0.239	0.591	100.000%
		Restricted Cubic Splines (Subjectively)	0.999	0.239	0.591	99.997%
	8 494	Linear Model	0.991	0.228	0.655	99.795%
		Categorized (Quartiles)	0.991	0.228	0.653	24.957%
		Categorized (Subjectively)	0.991	0.229	0.649	99.678%
		Quadratic Model	0.991	0.228	0.656	99.786%
		Fractional Polynomials	0.991	0.228	0.656	99.788%
		Restricted Cubic Splines (Data-Driven)	0.991	0.228	0.656	99.789%
		Restricted Cubic Splines (Subjectively)	0.991	0.228	0.656	99.791%

Abbreviations: C, C-statistic; RMSE, Root-Mean-Squared Error

3.2.4 Methods for detecting cumulative, protracted, time-lagged effects

The Distributed Lag Non-Linear Model (DLNM) discovered both the J-shaped relationship between absolute training load (sRPE) and injury probability (Figure 19D,H,L), and the linear relationship between relative training load (%Δ sRPE) and injury probability (*Paper III* Figure 4C,F,I), under all time-dependent scenarios. DLNM had the lowest mean external RMSE, internal RMSE and AIC, and narrowest 95% confidence intervals in all simulated scenarios, except in the Exponential Decay scenario for relative training load, where it had the lowest AIC, but the highest internal RMSE (Table 7).

The rolling average could model the constant scenario (Figure 19A), and EWMA could model the decay and exponential decay scenarios between absolute training load and injury risk (Figure 19F,J). No methods were able to detect the Direct, then inverse scenario (*Paper III* Figure S5). EWMA had the lowest mean external RMSE (aside for DLNM) under all scenarios (Table 7). REDI had the lowest performance across the board, with highest mean external RMSE, highest mean AIC, and it estimated that injury probability decreased (when the true probability increased) for each level of absolute training load under the exponential decay scenario (Figure 19K).

All methods of modelling absolute training load displayed poor coverage, ranging from 19% to 36% under all scenarios (Table 7), though the coverage estimates were uncertain (monte carlo standard error 0.6–0.9).

ACWR was not able to detect a relationship between relative training load (%Δ sRPE) and injury probability under the constant and exponential decay scenarios, while the week-to-week percentage change was not able to detect the relationship under the constant scenario. Both had broad confidence intervals and high internal RMSE and AIC compared with DLNM (Table 7).

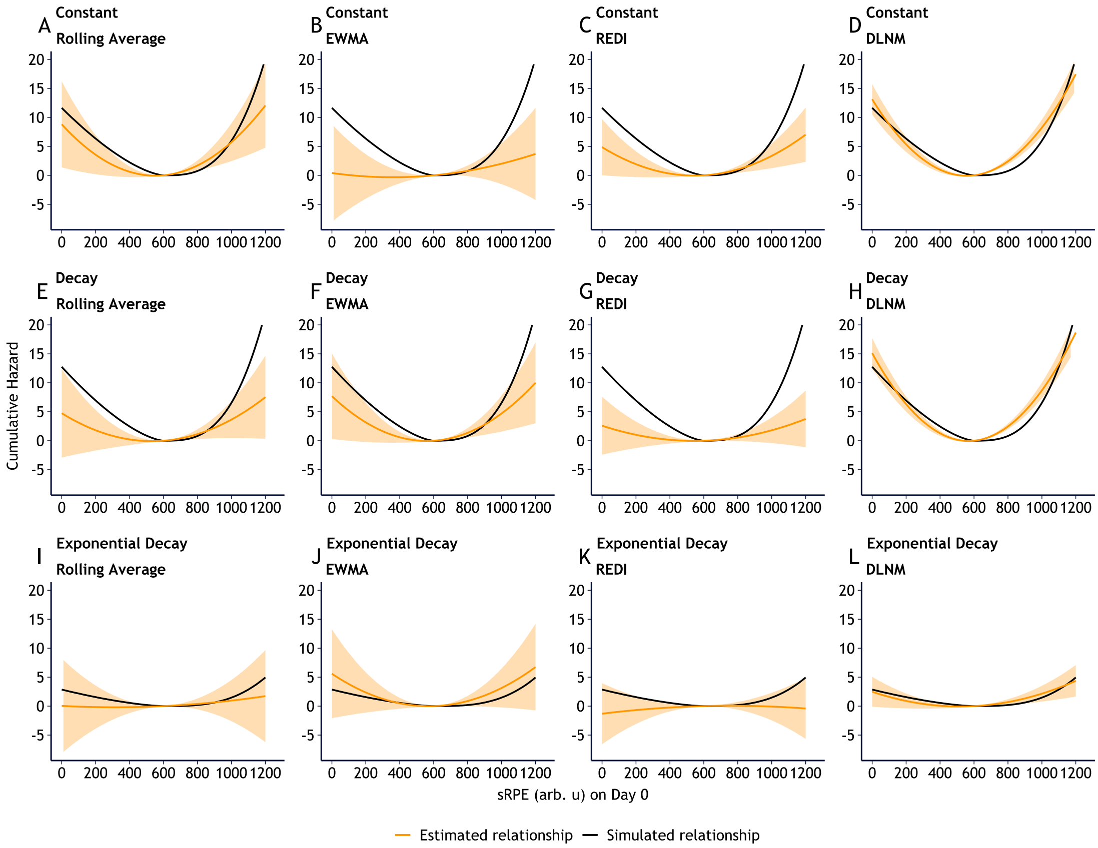

Figure 19. Risk of injury on the current day (Day 0) for each level of session Rating of Perceived Exertion (sRPE) as predicted by a Cox regression model run with four different methods for assessing training load. The black line is the simulated, true, relationship between sRPE and injury—the yellow line represents the ability of the method to model this relationship. Ideally, the yellow line should follow the black line as closely as possible for the highest accuracy. The Y-axis denotes the sum of all instantaneous risks of injury from the past up until the current day, the cumulative hazard. Relationships were simulated under different scenarios: (A–D) Constant: In the previous 27 days, the occurred sRPE contributed equally to injury risk on the current day; (E–H) Decay: The effect of sRPE was at its highest on the current day (Day 0) and reduced linearly for each day back in time; (I–L) Exponential Decay: The risk of sRPE was at its highest on the current day (Day 0) and reduced exponentially for each day back in time. The methods compared were: Rolling Average, the Exponential Weighted Moving Average (EWMA), The Robust Exponential Decreasing Index (REDI), and the Distributed Lag Non-Linear Model (DLNM). Yellow bands represent 95% confidence intervals. The figure shows 1 random simulation of 1 900 performed. Arb. u = arbitrary units.

Table 7. Mean performance of methods for handling time-dependent effects between training load and injury risk.

Relationship	Method	E-RMSE[1]	I-RMSE	AIC	Coverage	AW	Coverage MCSE
Absolute training load							
Constant	Rolling Average	4.85	0.113547	1422.9	35 %	5.175	0.90
	EWMA	4.77	0.113548	1423.4	36 %	5.172	0.91
	REDI	5.53	0.113557	1424.1	20 %	3.401	0.74
	DLNM	1.44	0.112434	1317.2	35 %	2.056	0.95
Decay	Rolling Average	5.38	0.113590	1421.8	30 %	5.169	0.87
	EWMA	5.17	0.113587	1421.9	32 %	5.126	0.88
	REDI	6.21	0.113605	1423.8	19 %	3.422	0.71
	DLNM	1.55	0.112245	1295.3	32 %	2.080	0.93
Exponential Decay	Rolling Average	2.13	0.113599	1424.7	85 %	5.547	0.58
	EWMA	1.88	0.113588	1423.9	85 %	5.371	0.61
	REDI	1.97	0.113603	1425.0	74 %	3.692	0.64
	DLNM	0.76	0.113368	1407.1	82 %	2.026	0.65
Relative training load (%Δ)[2]							
Constant	ACWR		0.113643	1426.2			
	Week-to-week %Δ		0.113646	1426.4			
	DLNM %Δ		0.113627	1389.3			
Decay	ACWR		0.113615	1424.7			
	Week-to-week %Δ		0.113617	1425.1			
	DLNM %Δ		0.113553	1383.5			
Exponential Decay	ACWR		0.113565	1423.3			
	Week-to-week %Δ		0.113566	1423.3			
	DLNM %Δ		0.113700	1401.4			

Abbreviations: ACWR = Acute:Chronic Workload Ratio; AIC = Akaike's Information Criterion; AW = Average Width of 95% confidence intervals; Coverage = Coverage of 95% confidence intervals; E-RMSE = External RMSE; I-RMSE = Internal RMSE; MCSE = Monte Carlo Standard Error; EWMA = Exponentially Weighted Moving Average; DLNM = Distributed Lag Non-Linear Model; REDI = Robust Exponential Decreasing Index; RMSE = Root-Mean-Squared Error

[1]Monte Carlo Standard Error for RMSE was < 0.001 for all 1 900 simulations.

[2]Due to differences in scale between methods and simulation for relative training load, external RMSE, coverage, and AW could not be calculated in a comparable manner.

3.3 Observed sports data results

None of the Norwegian football cohorts, when analyzed in *Paper II,* revealed any relationship between measures of training load and probability of injury (*Paper II* Supplementary Figure S5–S6). In *Paper IV*, the Norwegian elite U-19 model showed some signs of an association between 4-week chronic training load and injury risk (Figure 20B), though confidence intervals were broad for multiple spline-intervals, and in some cases, coefficients were inestimable (*Paper IV* Table S3). The model indicated the highest risk of injury if chronic load was low (sRPE), intermediate risk if chronic load was high, and lowest risk at medium levels of chronic load (Figure 20B). The slopes of acute load changed with different levels of chronic load, indicating an interaction between acute and chronic load.

In the elite youth handball players, a strong J-shaped relationship was found between sRPE and the probability of injury on the current day in a mixed effects logistic regression model ($p < 0.001$, *Paper II* Figure 2A). This relationship had the same shape in the frailty model used in *Paper III*, but with much broader confidence intervals ($p > 0.8$, *Paper III* Figure 5). An uncertain ∩-shaped relationship between sRPE (Day -4) and probability of injury in the next four days (Day-3 to Day 0) was found in *Paper II* ($p = 0.06$, *Paper II* Figure 2B). The DLNM in *Paper III* indicated increased risk of injury on the current day for high levels of sRPE sustained in the near past (Day -1 to Day-6), no effect of sRPE reported 6 days prior to the current day (Day -7), and thereafter (Day -8 to Day -27), high levels of sRPE indicated reduced risk (HR between 0.75 and 1.0 *Paper III* Figure 5).

The daily ACWR failed to adjust the numerator to the denominator (*Paper II* Figure S3), while the micro-cycle ACWR had no relationship in any models (*Paper II* Figure S5–S6).

The Qatar Stars League model showed reduced risk of injury with every minute in activity on the current day ($p < 0.001$, Figure 20A), which is expected when players end activity due to injury. Following the same pattern as the Norwegian elite U-19 model, highest risk was at zero and low chronic load, intermediate risk from high chronic load, and finally, the lowest risk was on days with medium chronic load, where multiple terms had significant p-values *(Paper IV* Table 1). The slopes of the relationship between acute load (minutes in activity on the current day) and injury risk varied considerably for different levels of chronic load, indicating an interaction. The risk declined rapidly for zero and low chronic loads, while it declined gradually for high and medium chronic loads (Figure 20A).

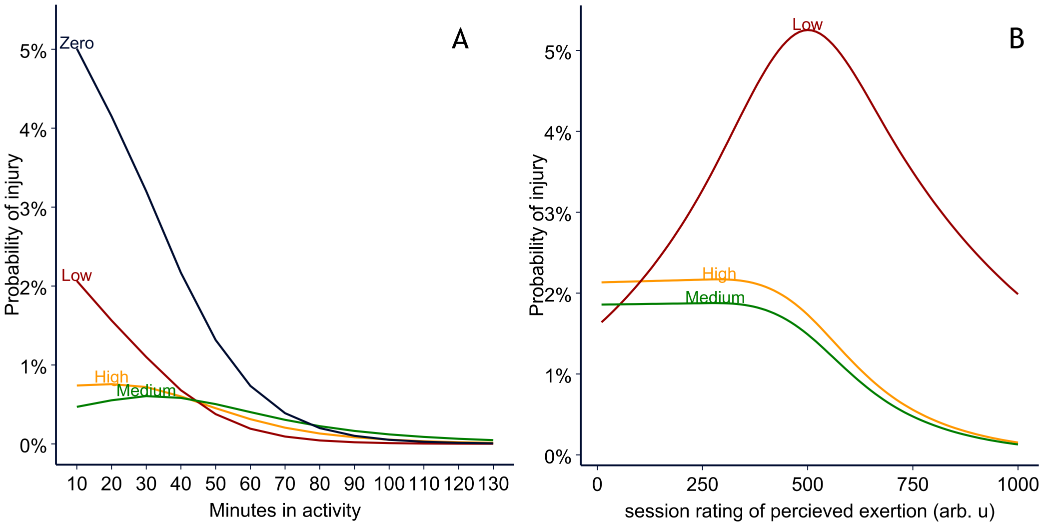

Figure 20. Probability of injury on the current day (acute load) for each level of training load variables, given different levels of cumulative chronic training load, predicted by logistic regression with an interaction term between acute and chronic load. Shown for (A) Qatar Stars League model (420 329 exposure values, 1 977 injuries), where training load was defined as the number of minutes in training and/or match activity, and (B) Norwegian elite U-19 model (4 719 exposure values, 60 injuries), where training load was defined as the session rating of perceived exertion, measured in arbitrary units (arb. u). For the Qatar Stars League model: zero = 27-day sum of 0 minutes, low = 180 minutes, medium = 1435 minutes, high = 1900 minutes. For the Norwegian elite u-19 model, low = 27-day sum of 80 sRPE (near zero), medium = 7 163, high = 8 800. See Paper IV Supplemental Table S1 for the exact chronic load profiles used.

4 Discussion

4.1 Missing data in training load

4.1.1 Missing data reporting practices in the field of training load and injury risk

The percentage of studies that reported whether or not they had missing data was alarmingly low (34%), especially compared to similar reviews in other fields, such as clinical RCTs (72% Díaz-Ordaz et al., 2014) and developmental psychology (57%, Jeličić et al., 2009). Of studies that reported whether they had missing data, only half (49%) reported the amount of missing training load data. The mean reported amount of missing data was 7.3%, which is surprisingly low for longitudinal studies (Karahalios et al., 2012). We speculate that this underreporting was under the mechanism of missing not at random; that the probability of a study reporting the amount of missing data depended on the amount of missing itself. We encourage future researchers to report missing data and how they were handled, and recommend reporting guidelines in Borg et al. (2022).

4.1.2 Sample sizes in the field of training load and injury risk

To our knowledge, *Paper I* was the first study to map the number of injuries assessed in training load and injury risk studies. By mathematical exploration, 96 participants with 48 events were needed for an accurate logistic regression model without independent variables (Riley et al., 2019). Logistic regression, which requires a binary outcome, is the most frequently used method in the training load and injury risk field (Windt et al., 2018). Although the ideal sample size will vary from study to study (Riley et al., 2019), the numbers seen currently are unlikely to satisfy minimum requirements (Figure 15). Consequently, it reduces the ability of studies to consider complexities such as within-subject correlations, non-linearity and confounding, along with increased risk of overfitting in prediction studies.

4.1.3 Handling missing data in training load measures

Imputation of missing training load observations may retain injuries that otherwise would have been removed and thus improve sample sizes. We recommend multiple imputation using predicted mean matching—if preferable, through collaboration with a sports biostatistician

(Sainani et al., 2021)—to impute missing data in continuous training load measures. In our simulation, PMM had improved performance over the other methods, with ≤ 3% bias up to 50% missing in sRPE under MCAR and MAR, and ≤ 5% bias up to 90% missing in total distance under MCAR, 50% under MAR. One can argue that this performance is no surprise, as it is a multiple imputation method. However, our secondary simulation showed that PMM sometimes had the lowest bias in the single imputation, and sometimes lowest in the multiple imputation framework, with no clear winner (*Paper I* Table S1); multiple imputation was developed mainly to improve the calculation of uncertainty estimates, standard errors and confidence intervals (Van Buuren, 2018).

We recommend imputing duration and RPE before calculating sRPE, if possible, as it had the lowest mean percent bias of the approaches to impute sRPE (1.4%).

PMM underperformed when imputing GPS variables if all the GPS variables were missing simultaneously. This indicates high correlations between the GPS variables. The only other information from which PMM predicted the missing GPS observations were match (yes/no) and the micro-cycle-day. PMM is prone to imputing duplicates if the predictors are poor or the non-missing sample size is too small (Van Buuren, 2018), which can explain why it performed worse than mean imputation and complete case analysis in this scenario. The aim of limiting the number of predictors in the imputation model was to allow complete reproducibility of the simulation, without violating anonymization laws. We argue, that if PMM had such superior performance to the other methods when GPS variables were present, and only declined when GPS variables were missing, it should perform even better in a real study. PMM was nearly on par with the other methods when sRPE was included in the imputation model. In our review of the training load and injury risk field, 15% of studies had both sRPE and GPS measures, and in a review of training load monitoring in football, 45% studies had both (Miguel et al., 2021). We encourage future researchers to include sRPE in the imputation model when imputing GPS variables.

If resources are not available, mean imputation might be considered in select circumstances. While the percentage missing should not be used to guide multiple imputation choices (Madley-Dowd et al., 2019), for mean imputation, the number of imputed observations relative to the observed observations may be a gauge of how much the imputation may bias the results. It can also be helpful to consider the distribution of the data. The total distance measure used in our simulation had a bimodal (camel-hump) distribution (*Paper I* Figure S1B), and the performance of mean imputation, whether by the weekly mean or the player mean, varied between being among the worst, to among the best (*Paper I* Figure 6–7), and it was particularly volatile under MAR (*Paper I* Figure 7). Weekly means, as opposed to player means, had the best performance in 4 of 6 scenarios of total distance under MAR, which supports results and recommendations in Benson, Stilling, et al. (2021), but it may have been due to chance. The performance of the two mean imputation variants in each of the 1 900 simulations may have been strongly influenced by which observations were missing, determining whether the mean would be representative of the training load or not.

Plotting the imputed versus the observed values, such as in *Paper I* Figure S2–5, may also aid in determining how much the mean imputation will introduce bias. It could also be worthwhile to check how many injuries would be discarded with the complete case analysis approach, as it generally had improved performance over mean imputation, especially for sRPE.

For count data (e.g. number of strength training sessions), PMM can also be a valid approach if data are not extremely skewed (Van Buuren, 2018). In *Paper IV*, we compared imputation performance from a Poisson regression imputation and PMM before settling on PMM for the number of minutes in activity in the Qatar Stars League data. Multiple imputation by the random hot deck approach is also a promising alternative for count data if constraints are present (C. S. Wang, Tyrel et al., 2020), such as, if total number of training sessions is three, than the number of strength training sessions cannot exceed three. In the random hot deck approach, this can be specified.

In our discussion and recommendations, we assumed that training load data are MAR. Using methods of multiple imputation handles this missing mechanism, and when using such methods, tests for missing completely at random are generally not necessary. Longitudinal data are generally MAR, especially participant-reported data (Barnett et al., 2017), and Benson, Stilling, et al. (2021) reported their sRPE data to follow a pattern of MAR. In all football populations studied in this dissertation, the data were MAR from various causes (*Paper I* Figure S2, *Paper IV* Figure S2). Note that exploring the missing data pattern with visualizations is still useful to understand which variables should be included or excluded from the imputation model, among other considerations(Van Buuren, 2018).

In contrast to the MAR assumptions of the football cohorts, the Norwegian elite youth handball data studied in *Paper II–III* was suspected to be MNAR (Bjørndal et al., 2021), with the theory that players who were more invested in the sport—and thus also trained more—had a higher probability of responding. The population had high levels of training load, which may be the result of selection bias. Although using imputation methods, even methods that handle MAR, may not necessarily be valid, complete case analysis will introduce selection bias, as demonstrated in our simulation (Figure 16, Figure 17). In conclusion, if MNAR is suspected and the true observations cannot be gathered retrospectively, it can still be beneficial to impute with PMM to conserve injuries, but the implications should be considered and discussed.

4.2 Non-linearity between training load and injury risk

4.2.1 Non-linearity in training load and injury risk relationships

Two main theories suggest non-linearity between training load and injury risk: (1) both too little and too much absolute training load increases risk, and reduced risk is at intermediate levels (Gamble, 2013), (2) contribution of training load on injury risk decays exponentially with time (Williams, West, et al., 2017). Note, theory 1 pertains to non-linearity in chronic (past) training load only, and theory 2 pertains to non-linearity between training load, injury risk, and time.

We speculated that the football cohorts studied in *Paper II* showed no association between sRPE and injury risk due to limited sample sizes; most of the statistical models had fewer than 50 events (*Paper II* Supplementary Table S2). When the Norwegian elite U-19 football was later studied again in *Paper IV*, it indicated, with high uncertainty, non-linearity between sRPE on the current day (acute load) and injury risk, and between 27-day sRPE (chronic load) and injury risk. The risk for each level of acute load decayed exponentially, but only for medium and high chronic load; for low chronic load, the risk first increased and then decreased in a sharp parabola.

Interestingly, the Qatar Stars League (QSL) model had a similar risk pattern between minutes in activity and injury risk, but with narrow confidence intervals and multiple significant p-values ($p < 0.05$). In this model, the risk decayed exponentially for each level of acute load if chronic load was zero or low, and decayed gradually if chronic load was medium or high. In essence, both the QSL and the Norwegian model suggested a substantially elevated risk at low chronic loads and decreased risk at medium chronic load, while high chronic load was at intermediate risk. This supports the theory that both too little and too much training load (in the past) may increase risk of injury compared with medium levels (Gamble, 2013; Windt & Gabbett, 2017).

In the models explored with DLNM, that is, the Norwegian elite youth handball model in *Paper III* and QSL model in *Paper IV*, effect sizes of absolute chronic load were exponentially smaller for each day in the past. This supports the theory that the contribution of training load on injury risk decays exponentially with time (Williams, West, et al., 2017).

In the QSL model (*Paper IV*), the U-shaped pattern of increased risk with high and low levels of training load was present every day in the past from day -1 to day -27. This pattern was not present in the handball model (*Paper III*). The model suggested, with high uncertainty ($p >= 0.8$), that high levels of training load increased risk on the current day, but *decreased* risk if they occurred approximately ten days ago or more. This is more supportive of the theory that acute (current) training load increases risk and chronic (past) training load decreases risk (Gabbett, 2016). This theory has been critiqued severely (Franco M Impellizzeri et al., 2020), however, and the model may fit by coincidence. The sparse data in high training load levels likely caused imprecision, and the large amounts of missing data and poor timeliness may have introduced more noise and uncertainty. These results indicate that the relationship is either too complex, or has too small effect sizes, for data of this size (471 injuries) and quality to be sufficient for this population. Notably, the sample size was larger than most training load and injury risk studies reviewed in *Paper I* (median = 85 injuries). To explore non-linearity, whether in training load's relationship with time, or inherently, data should be collected with this goal in mind.

4.2.2 Handling non-linearity between training load and injury risk

Fractional polynomials and Restricted Cubic Splines (RCS) could accurately detect non-linear relationships between training load and injury risk and had superior performance to the other methods in our simulation. This reproduces the results in Carey et al. (2018) and substantiates their recommendations.

In our results, the performance of restricted cubic splines depended on the location of the knots, from being among the best methods to among the worst. This contradicts previous claims that the number of knots is more important than the location (Harrell Jr, 2017; Stone, 1986). We believe this is context-specific, and in our case, where the data was exceedingly skewed, data-driven location biased results. Indeed, using a data-driven approach for knot-placement of RCS may have biased results on the observed data analyses in *Paper II* and *Paper III*. The Norwegian elite U-19 football model went from showing no relationship between sRPE and injury on the current day in *Paper II* to a small effect estimate in *Paper IV*—though still with confidence intervals overlapping with 1. Knot locations may have contributed to this change.

Collinearity was an issue in both the Norwegian elite U-19 model and the Qatar Stars League football models in *Paper IV*. It is likely that some spline terms were too correlated, and consequently, confidence intervals surrounding predicted probabilities could not be estimated. We chose splines to more readily detect sudden changes in the direction of effect, but I reason that the collinearity issue might have been amended by using fractional polynomials (FP). If splines knots are located too closely, the model may struggle to estimate the difference between two very similar intervals. In contrast, FP fits a function on the entire range of values, and therefore, collinearity between intervals within the range should be of no concern. This property also makes it less susceptible to bias from data-sparse regions (Binder et al., 2013). FP also had lower RMSE than both RCS-versions for the J-shaped relationship and better coverage in all simulated scenarios in *Paper II*. Since the splines-function formed a parabola in the relationship between acute load and injury risk, we could also have chosen a quadratic model for simplicity and interpretability in this specific case.

Recently, thin plate splines have been used to study training load and injury (Wang et al., 2022). Thin plate splines is a two-dimensional application of the cubic splines, meaning it can determine the combined effect of two continuous explanatory variables on an outcome. This may be suitable for handling multiple measures of the same training load dimension, such as total run distance and sprint distance. On the other hand, it is a more advanced method that may require collaboration with a sports biostatistician (Sainani et al., 2021).

By using fractional polynomials or RCS (methods of model specification), step 4 in Figure 6, categorization, can be eliminated. The performance of categorization was worse in our study than in Carey et al. (2018), as categorization by quartiles could not even accurately model the linear shape (coverage of 95% prediction intervals = 25%, vs. >99% for all other methods). Categorization increases both Type I and Type II errors rates (Harrell Jr, 2017), assumes that observations within intervals are equal, and assumes that the relationship shape between training load and injury risk is flat within intervals (Figure 18)—both of which may be a more unrealistic assumption than linearity (Frøslie et al., 2010). It also reduces comparability and reproducibility (Holländer et al., 2004); reduces precision and power (Collins et al., 2016), thus, requires larger sample sizes, and requires more hypothesis tests, thereby risking multiple testing issues (Dalen-Lorentsen, Andersen, et al., 2021).

Categorization has such severe disadvantages (Frøslie et al., 2010), it can only be justified in a study where analysts have no choice; where the training load measures were pre-collected as categories. I do not believe that results from categorization can be considered evidence of, or refute, a non-linear, U-shaped relationship between training load and injury risk, as some studies have claimed (Cross et al., 2016; Malone et al., 2018; S. Malone et al., 2017; Sedeaud et al., 2020); at best, such results are hypothesis-generating.

If categories are of interest, I urge researchers to model training load as a continuous variable regardless, and rather show the predicted injury counts or probabilities for each category of interest post-analysis, as recommended in (Rhon et al., 2022).

4.2.3 Non-linearity between predictors of training load

When simulating a relationship between training load and injury and introducing missing data with different mechanisms in *Paper I*, we did not interfere with the relationship between other variables in the observed football data. Continuous variables, particularly age and time (e.g. calendar week), may have a non-linear relationship with training load (Gabbett, 2016). Given our results in *Paper II*, the imputation model for regression imputation and PMM may also be improved by specifying fractional polynomials or RCS for such predictors in future studies. Classification and regression trees, like random forest, and unsupervised learning, like K-means clustering, may also be useful in imputation of missing training load data with non-linear relationship, as overfitting is not as problematic when the classification is not meant for external use (Harrell Jr, 2017).

4.3 Time-dependent effects in the training load and injury risk relationship

4.3.1 Modelling time-dependent effects with distributed lag non-linear models

The distributed lag non-linear model (DLNM) had an impressive level of performance for estimating the cumulative, protracted, time-lagged effects of training load on injury probability in the simulation study of *Paper III*. It had the highest accuracy, precision and certainty in all simulated scenarios, and since it was the only method that uncovered most of the non-linear shapes between training load, injury probability and time, it was also the most flexible. Finally, the DLNM does not require partitioning the data in time intervals before analysis, and using this method eliminates step 2 in Figure 6. We promote the DLNM method as the current best practice for estimating the relationship between cumulative, long-term training load and injury risk.

While DLNM was ideal in a simulation, in practice, it had a few practical considerations that should be discussed. First, it has limited software implementation. The DLNM R-package was developed in the field of environmental epidemiology, and is compatible with most types of models, including mixed models and time-to-event analysis. It is not necessarily compatible with certain other R packages used to handle other, more edge case complexities. I have experienced setbacks using the DLNM with multistate modelling, which may be relevant for studies of gradual onset injuries.

Second, in the context of training load, it is often necessary to remove days in which the athlete is not at risk, i.e., is not participating in any activity and therefore has no training load. Days in which the athlete did not participate in activity is still relevant for the cumulative, long-term training load, and should not be removed from the past observations in the DLNM-calculation. The DLNM R package had no option to handle this automatically, and although we solved it manually, confidence intervals and p-values for the chronic load coefficient estimates were based on the smaller sample size of the acute load, meaning that uncertainty estimates were larger than in reality. In addition, DLNM R-package prediction functionalities could not be used, which hampered our analyses in *Paper IV*.

Finally, to explore time-lagged effects with DLNM, training load and injury measures must be collected and analysed at the daily level. This may not be feasible for many studies.

4.3.2 Other alternatives to handling time-dependent effects

Although DLNM has some practical considerations, we implore researchers to consider this method over other options currently available, even if it entails collaboration with a biostatistician (Casals & Finch, 2017; Sainani et al., 2021).

Menaspà (2017) illustrated how weekly training load patterns become washed out with rolling averages, to which Drew et al. (2017) asked for evidence of a better alternative, and Williams, West, et al. (2017) responded with the EWMA. The EWMA was later considered the improved option because relative training load had larger effect sizes in its association with injury risk when ACWR was calculated by EWMA than by the rolling average, in Australian football and elite rugby union (Murray et al., 2017; West et al., 2020). Although the same comparison was done in female football and recreational running, the opposite result was found (Nakaoka et al., 2021; Sedeaud et al., 2020). Sedeaud et al. (2020) additionally compared ACWR by REDI (Moussa et al., 2019), which had one more significant p-value than ACWR by rolling average. The results in Sedeaud et al. (2020) were inconsistent; the methods did not agree which ACWR zones were significant. This disagreement, by our results in *Paper II* and Carey et al. (2018), may have been exacerbated by categorizing ACWR.

I argue that the results in Murray et al. (2017), S. W. West et al. (2021) and Sedeaud et al. (2020) are not enough to declare EWMA or REDI an improvement over the other, nor over the rolling average. Either method may have had Type I or Type II errors without knowing. In a simulation, the true relationship is known, and the ability of methods to detect relationships is more easily assessed. In our simulation, all three methods of aggregation had poor performance relative to DLNM. While they could model the simulated relationships somewhat (Figure 19), DLNM had ca. 1000 points lower mean internal RMSE than all three for the constant and decay relationships, and 300 lower for the exponential decay (*Paper III* Table 1).

REDI consistently had the worst performance (highest RMSE and AIC), and even modelled an inverse relationship between sRPE and injury probability in the exponential decay scenario (advocated as the most likely scenario (Williams, West, et al., 2017)), when the true relationship was the opposite. Yet, it had narrower confidence intervals on average (covering on average 3 cumulative hazard) than EWMA and rolling average (5 cumulative hazard), meaning it was precise in its bias. This can mislead well-intentioned analysts into believing REDI is the better choice in a real study, when it shows narrower confidence intervals and perhaps significant p-values where EWMA or rolling average does not. This might have happened in Sedeaud et al. (2020).

Renfree et al. (2021) hypothesized that the effect of training load levels depends on how the training load is distributed across a week, and with methods of aggregation, this information is lost. EWMA and the rolling average had much broader confidence intervals (average 95% CI width (AW) = 5) than DLNM (AW = 2) in *Paper III*. I suspect methods of aggregation may increase risk of Type II errors. In a study on the New York City Marathon (Toresdahl et al., 2022), 30 independent tests were run with different combinations of time period and training load measures. Given the multiple testing combined with a high sample size (699 runners), one would expect multiple significant p-values. Only two tests had significant results, and that aggregation combined with categorization, both of which reduces power (Frøslie et al., 2010), may have led to negative findings.

The EWMA performed better than the rolling average (lower RMSE and AIC) when we assumed that the contribution of training load decayed linearly or exponentially. This scenario is more realistic than the assumption that training load contributes equally to injury risk regardless of time since the activity. I therefore recommend the EWMA over the rolling average.

Lazarus et al. (2017) suggested a modification of the EWMA in studies on performance, which later Coyne et al. (2022) marketed as equal to the traditional EWMA[4]. We have no reason to believe that the modified EWMA would perform better than the traditional EWMA. It weighs recent observations less—so that it is more similar to the rolling average, and thus, forms the assumption that the relationship decay is more constant than the traditional EWMA. More research is needed with explorative methods like DLNM to determine whether this assumption is reasonable, before it can be advocated.

When we attempted to use EWMA in *Paper IV*, it failed to discover a relationship between chronic load and injury risk in the Qatar Stars League data consisting of 1 136 223 training load observations and 1 977 injuries. Given such a large sample size, we speculate whether EWMA could have found the relationship that DLNM found at all, even if $n \to \infty$. After all, EWMA tries to describe 28 data points with one number; an issue shared between all methods of aggregation. While we do not consider EWMA to necessarily be a wrong choice for a training load and injury risk study, and consider it an improvement over the rolling average, researchers should be aware of the increased uncertainty and need of a larger sample size with this method, especially compared with DLNM.

4.4 The effect of relative training load on injury risk

4.4.1 Absolute versus relative training load

In addition to the dimensions of external and internal training load, training load can be divided into absolute training load, absolute change in training load, and the relative (change in) training load. This distinction is important, as hypotheses of how absolute training load affects injury are

[4] Traditional refers to the EWMA calculation in Williams, West, et al. (2017), used in this dissertation.

different relative training load (Gabbett et al., 2016; Gamble, 2013), and may have different causal pathways to injury. Absolute training load describes how much the athlete was exposed to training load, and may answer “How much is too much?”, and whether there are thresholds of too little training (Gamble, 2013). Relative training load describes, in theory, the change between current training load and the amount of training load the athlete is accustomed to (Gabbett, 2016). This measure may answer how much change in training load is too abrupt a change for tissue to tolerate (Vanrenterghem et al., 2017).

Studies that measure training load by the ACWR (Andrade et al., 2020), assess relative training load only.[5] Often, the type of training load is not specified, including in methodological studies that compare methods of assessing absolute training load with methods of assessing relative training load as though they are the same (Coyne et al., 2021; Moussa et al., 2019). The effect estimates for a method of absolute training load (such as EWMA) describes different training load phenomena than that of relative training load (such as ACWR), and may be larger or smaller based on the scale (Impellizzeri et al., 2021), not necessarily because absolute training load contributes less to injury risk.

In *Paper III*, we made a simulation to compare methods of absolute and relative training load separately. We experienced that the two constructs required different assumptions, and had different challenges. To answer the question “At what timepoint in the past does the effect of the magnitude of (absolute or relative) training load change?”, DLNM could be used to explore such time-dependent effects in absolute training load, but in relative training load, this had to be calculated and assumed before-hand. In addition, training loads of 0 were treated like any other value for absolute training load, but was a challenge when assessing relative training load, where either the numerator or denominator in a ratio could be 0.

Sedeaud et al. (2020) compared the performance of EWMA and REDI with the traditional, rolling average ACWR by including them within the same ACWR calculation. Scientists should follow their lead by ensuring methods are comparable, and be clear on whether methods aim to capture absolute, relative, or other training load constructs.

[5] Although the coupled ACWR is not a true measure of *change* in training load (Wang et al. 2020), it is still, from a statistical standpoint, a measure of relative as opposed to absolute training load

4.4.2 How to estimate the effect of relative training load on injury risk

The acute:chronic workload ratio, a method for assessing relative training load, has been a hot topic in the field of training load and injury risk the last decade (Figure 1). Discussions surrounded the advantages and disadvantages of the ratio (Blanch & Gabbett, 2016), and how the ratio should be calculated (Lolli et al., 2019). One issue was how to determine the acute and chronic time intervals, as calendar weeks may be arbitrary for many sports (Franco M Impellizzeri et al., 2020). In *Paper II*, we tried the micro-cycle uncoupled ACWR. Since the calculation moved iteratively from one micro-cycle to the next (as opposed to from one day to the next), and multiple injuries sustained in the period was considered 1 event, the sample size of injury events was reduced. While this is an improvement over the traditional one week to the next, it still reduces statistical power. In addition, since time intervals are of unequal size, one value to the next may not be comparable: Is one micro-cycle ACWR larger than the other because intense activities were sustained during a handful of days, or because less intense activities were spread across a larger number of days? This difference in distribution may be important to capture when assessing injury risk (Renfree et al., 2021). A potential solution may be to include the micro-cycle day as a variable in the statistical model alongside the daily training load values, such that confounding effects from the type of day is accounted for—if relevant to the sport and research question. This was successfully done in *Paper I* to improve prediction of training load values used to impute missing data.

In *Paper III*, the daily coupled ACWR was compared with other methods in a simulation, where it failed to detect the simulated relationship between relative training load and injury risk under the constant and exponential decay scenarios (*Paper III* Figure 4). The week-to-week percentage change was not much better, though, as it was also unable to detect the constant scenario. In addition, which of the two had the lowest internal RMSE and AIC varied between scenarios, with negligible differences. Replacing the ACWR with the week-to-week %-change is not necessarily an improvement.

Although the simulation confirmed concerns of the ACWR (and perhaps the week-to-week %-change), it was still difficult to determine how relative training load should be assessed. The DLNM had superior performance overall, but the simulation assumed that daily percentage changes affected injury risk, and these must be calculated before running DLNM. Using this method, analysts would still have to subjectively determine time interval cut-offs for percentage changes, and consider how to deal with numerators or denominators of 0.

Discussion

In *Paper IV*, we attempted to solve this issue by modelling acute and chronic loads separately, as proposed by C. Wang et al. (2020). Here, the chronic load (Day -1 to Day -27) was modelled with the DLNM without any calculations of percentage change before-hand, but assessed as the absolute chronic load. Only the current day (Day 0) was considered the acute load. We argued that the current day was sufficiently different from past training load days, given hypotheses and assumptions, to warrant a separate variable. No ratio was calculated; the idea was that the injury risk of acute load, dependent on the level of chronic load—and vice versa—can be discerned by including both in the same model. Our results showed the effect of acute load for different levels of chronic load, in both the Qatar Stars League football population and the Norwegian elite U-19 population. Using such methodology has several advantages:

(1) The acute load is properly adjusted for chronic load. The daily uncoupled ACWR calculated in *Paper II* is an example where the ratio failed to adjust numerator to the denominator (*Paper II* supplementary Figure S3), and this could have been avoided.
(2) Likewise, the chronic load is properly adjusted for the acute load. In many cases, the chronic load is of more interest than the acute load, as indicated by studies that analyze the chronic load for different levels of ACWR (such as in Bowen et al., 2020; Stares et al., 2018).
(3) The model outputs effect estimates for acute load and chronic load separately. Unlike the ACWR, which describes both with a single number, researchers can determine which is more important regarding injury risk (C. Wang et al., 2020).
(4) Training load values of 0 are modelled in the same manner as other values. If researchers would otherwise have removed chronic loads of 0, this also reduces missing data and selection bias.
(5) Small absolute changes in training load can correspond to extremely large relative changes if the denominator (chronic load) is small. For instance, being a cricket bowler, going from throwing 5 balls a week to 15 balls a week is a 300% relative increase. Such relative increases were at times not considered clinically meaningful, and typically also removed from analyses (Hulin et al., 2016; Stares et al., 2018). By modelling acute and chronic loads separately, this problem is solved.
(6) By use of DLNM on the chronic load, time-dependent effects can be explored without aggregation, and without partitioning the data in potentially arbitrary time intervals.

Other than practical considerations of the DLNM, the "separating the acute from the chronic" method has only one major disadvantage. By only considering the current day as the acute load, it assumes that there are no effects of relative training load in observations further in the past. For instance, if training load sustained 5 days in the past, relative to that sustained 10 days in the past, increases the risk of injury on Day 0, this is not assessed. Such relationships of relative training load within the long-term training load may be important features in prediction studies. In causal inference, the consequences of ignoring such relative relationships depends on whether they form confounding pathways on the training load dimensions of interest. For example, if a high training load one day in the past (Day -1), relative to that sustained 2 days in the past (Day -2), affects the absolute training load level sustained on the current day (Day 0), and this *also* affect the injury risk on the current day, it forms a confounding pathway for the risk assessment of acute load (Day 0) on injury risk. This can be relevant in real-world sports settings where coaches modify training schedules after "spikes" of training load. Notably, the ACWR also has this challenge, as it, too, only assesses one time period relative to another.

While modelling acute and chronic separately is a valid approach to achieve statistical effect estimates, I caution that it is only one of several methodology components needed to approach an unbiased estimate of causal effects. Causal inference requires study design and/or methodology that accounts for confounding and other sources of bias (Stovitz & Shrier, 2019).

4.4.3 Interaction between acute and chronic training load

Bittencourt et al. (2016) described how multiple factors or causes of injury may contribute to injury risk multiplicatively rather than additively, and recommended researchers in sports medicine consider such interactions to capture the full extent of the sports injury problem. In causal inference, identifying interactions can determine mediating and modifying effects which can, in turn, improve injury prevention strategies.

We discovered an interaction between acute load, defined as the current day of activity, and chronic load, defined as the previous 27 days of activity, in their relationship with injury risk in Qatar Stars League football players (*Paper IV*). Tendencies of a similar interaction were found in the Norwegian elite U-19 model, albeit with higher levels of uncertainty (some spline terms were inestimable, and confidence intervals were broad). Both analyses showed a pattern of steeper slopes (decreasing fast) per level of acute load if chronic load was low, but gentle slopes (decreasing slowly) if chronic load was medium or high (Figure 20). Since we only searched for associations, and did not apply a causal inference approach, we cannot with any certainty give the patterns in the data a clinical explanation. Nevertheless, these results demonstrate that this simple approach can be used to test the too much, too soon theory (Gabbett et al., 2016), given that other methodological considerations for causal inference have been addressed. It can also be used to investigate which training load construct contributes most to injury risk: changes in training load or the absolute magnitude of training load.

Some studies have investigated interactions by assessing the association of acute load or ACWR with injury risk for different subgroups of chronic load (Bowen et al., 2020; Hulin et al., 2016; F. Impellizzeri et al., 2020). Bowen et al. (2020) found increased risks of non-contact time-loss injuries in football if uncoupled ACWR was ≥ 2 while the previous 4-week rolling averages (chronic load) of various GPS measures were defined as low. The pattern was present in total distance, low-intensity speed distance (distance covered below 14.4 km/h), accelerations and deaccelerations, with odds ratios between 4 and 6 and $p < 0.05$, based on 91 injuries. This aligns well with our result of increased injury risk at low and zero chronic loads; however, for acute load, comparability is low, as the authors defined acute load as the previous 1 week, including the current day, while we only considered the current day as the acute load. In addition, we used minutes in activity and sRPE as training load measures, while they used GPS measures.

The definition of low chronic load in Bowen et al. (2020) was any observation below the median of each GPS measure. GPS measures are often left-skewed (Thoseby et al., 2022), but they may also have a bimodal distribution if training and match loads have considerably different distributions, which was the case for the total distance measure in *Paper I* Figure S1B. This means that it is unclear whether the "low" chronic load category covered running distances from 0 to moderate levels, or perhaps covered running distances representative of typical training sessions. In addition, due to sparse data in the subgroup, the authors decided not to assess the association between ACWR and injury while chronic load was defined as high (above the median), which was a good choice as the results would be highly uncertain, though this meant interactions could not be determined. The advantage of modeling interactions, rather than performing subgroup analyses, is an efficient use of the available data (Brankovic et al., 2019). Modeling interactions also avoids categorizing the acute or chronic loads, a requirement of subgroup analyses that reduces statistical power further (Collins et al., 2016).

Stares et al. (2018) studied the association between ACWR and 133 non-contact time-loss injuries in 70 Australian football players. Commendably, they used generalized estimating equations with a Poisson link, fitted the ACWR of each GPS measure and sRPE with a quadratic term (independently), and modelled their interaction with four categories of chronic load. This allowed them to account for within-individual correlations and avoid the many pitfalls of categorizing ACWR described in section 4.2.2. At low levels of sprint distance ACWR, the two categories of intermediate chronic loads (low and high) were at lowest risk. However, the slopes changed as ACWR increased, and at high ACWR levels, the two categories of extreme chronic load (very high and very low) were at lowest risk (Stares et al., 2018, Figure 1B). Since the level of acute load is obfuscated by the ACWR calculation, and low ACWR may mean that players were injured early in the week and thus had reduced loads the remaining days of that week, the results in Stares et al. (2018) are difficult to interpret. In addition, they removed extreme ACWR values resultant of regular training after periods of extremely low chronic loads, which means the lowest chronic loads were not assessed. Nevertheless, the overall impression of the results in Stares et al. (2018) is that an interaction between acute and chronic loads may also be present in Australian football; however, further research involving interactions is needed for more evidence.

Interactions should only be included in studies of causal inference, however, if there are hypotheses or rationale that suggest there may be an interaction. Searching for interactions in a data-driven manner increases risk of Type I errors and spurious correlations (Harrell Jr, 2017). Interaction terms require higher sample sizes, and if they are not required, a simple model may be a more efficient use of the data. Lathlean et al. (2022) and Ramskov et al. (2021) are exemplary, as the authors justified the testing of interactions with plausible theories.

4.5 Causal inference versus prediction modelling

Sampson et al. (2017), in response to the methodology editorial chain of 2017 (Drew et al., 2017; Menaspà, 2017; Williams, West, et al., 2017), argued that some methods may be more appropriate for some sports and populations than others, and one model is not necessarily the best fit for all. They recommended applying multiple methods on the same data across different sports to determine the best method(s) for each of them. Recently, Coyne et al. (2022) recommended data-driven approaches (including the AIC) to determine the best metric (e.g. EWMA vs. rolling average), the length of acute and chronic time periods, and variable selection, in monitoring training load for performance.

Data-driven practices appear to be prevalent also in the field of injury risk (Franco M. Impellizzeri et al., 2020b), where researchers assess multiple metrics of absolute and relative training load (e.g. including both ACWR and EWMA), calculation variants (e.g. coupled and uncoupled ACWR), time intervals (e.g. 3-day acute load and 7-day acute load), categories (e.g. high vs. low training load and medium vs. low), variables (e.g. total distance and high speed running distance), and injury definitions (e.g. contact and non-contact injuries), as illustrated in Figure 6. The current practices result in multiple testing issues (Dalen-Lorentsen, Andersen, et al., 2021), and interpretation becomes nearly impossible. Why did the 3-day acute period, 28 day chronic period, coupled ACWR by EWMA show up as significant, when the *uncoupled* ACWR by EWMA did not (West et al., 2020)? Inconsistent results perplex researchers and practitioners (Franco M. Impellizzeri et al., 2020b), and muddle systematic reviews (Griffin et al., 2020).

These data-driven approaches to search for the most optimal variables and metrics—as decided by significance or predictive ability—may stem from methodology used in studies of prediction. Data scientists regularly run different machine learning models (algorithms) and choose the one with the best predictive ability (e.g. Jamil et al., 2021). Deciding the number of trees in a random forest, or the number of knots in splines, is done through data-driven approaches in so-called hyperparameter optimization (Majumdar et al., 2022). Regularization methods like lasso or elastic net are used to peal away uninformative predictors (Zumeta-Olaskoaga et al., 2021). Researchers in prediction studies may remove variables for many reasons. If variables are time-consuming or expensive to collect, it would be easier to implement the prediction model if it can predict sufficiently without them. In high-dimensional data with no apriori information to determine predictors, such as in genome studies, identifying the most important predictors can be hypothesis-generating (Johnstone & Titterington, 2009). When uninformative variables offer more noise than signal, they may worsen predictions (Han et al., 2008). Lastly, perhaps the sample size is too small to justify using all predictors available (Harrell Jr, 2017). The latter is the only one that applies to training load and injury risk studies of causal inference.

In causal inference, the scientist aims to achieve an unbiased estimate of the effect of the exposure of interest on the outcome. The variable(s) that describe the exposure of interest are included in the model (or test) to determine the effect size of each. Other variables are only included to eliminate or reduce confounding, or perhaps, to understand how much of the effect is mediated by other constructs (such as in Lathlean et al., 2022). Including all variables available, without drawing assumptions of how they play into the causal pathway between training load and injury, risks Table 2 fallacy—interpreting all coefficients in the model as though they describe the total effect of each construct, when in reality, they do not (Westreich & Greenland, 2013). Data-driven approaches for variable selection are therefore not used. If a training load variable offers more noise than signal (implying that it does not affect injury risk) this is considered interesting, and whether it reduces the collective predictive ability of the model is not important. These are fundamental differences between studies of prediction and studies of causal inference.

In *Paper II*, a linear model had superb predictive ability for the non-linear shapes (C-statistic > 0.8), even though it could not model the relationship at all, because the training load values were skewed. Such a model could potentially still be used in a prediction study (especially if external validity is of no concern), but in causal inference, it would be disastrous. The wrong conclusions would be drawn on the relationship between training load and injury risk, and in turn, the wrong recommendations for injury prevention would be made.

Using p-values to dictate the inclusion of a variable or metric, such as backwards stepwise elimination, is unadvised for both causal inference and prediction (Derksen & Keselman, 1992; Steyerberg et al., 1999). Even though variables are not significant, they can still be informative predictors, and even though they are significant, they may still be biased. In our simulation in *Paper III*, the REDI had poor accuracy, but higher precision than the other methods, and could potentially bias assessments of both effect size and predictive ability.

I recommend choosing the method(s) that best fits the assumptions, resources and aims of the study, and not fall into the temptation of including them all. I agree with Sampson et al. (2017) that some methods may be better suited in some sports and populations, but, the assumptions of said sport should be considered when choosing methods, and not be up to data-driven approaches influenced by chance to decide (Gamble et al., 2020), especially for causal inference.

4.6 Machine learning alternatives

Recently, machine learning has been recommended for sports injury research (Bittencourt et al., 2016; Nielsen, Shrier, et al., 2020; Ruddy et al., 2019). While machine learning methods may be used for other aims, in this discussion, we will focus on prediction, where they are more often used (Claudino et al., 2019).

One machine learning branch that has been recommended, are classification methods. So-called Classification And Regression Trees (CARTs) classify data in binary groups, such as "injured" and "not injured". Examples include decision trees, random forest, and support vector machines (Sidey-Gibbons & Sidey-Gibbons, 2019), and are the most used machine learning methods so far in the training load and injury risk field (de Leeuw et al., 2022; Lövdal et al., 2020; Majumdar et al., 2022; Mandorino, Figueiredo, Cima, et al., 2022; Nunes et al., 2022). They are run with minimal to no model specification input from the user with minimal assumptions, and can—in theory—predict outcomes from associations in complex non-additive systems, including contexts of non-linearity, time-dependence and interactions (Bittencourt et al., 2016).

CART algorithms are a series of if-else statements that are used to predict the classification of the observation. Through recursive partitioning, predictor(s) with the best-performing binary split are chosen. For continuous predictors, such as training load, this split must be above and below a set cut-off; it dichotomizes the predictor. While these binary categories may be more parsimonious than our categorization approaches in *Paper II*, as they are calibrated through predictive ability, they still involve the strong assumptions of dichotomization—that the relationship between predictors and outcome are flat within intervals (Royston et al., 2006). Rhon et al. (2022) demonstrated lower predictive ability of injury prediction models when predictors were dichotomized rather than modelled as continuous.

CARTs are also more prone to overfitting than regression analyses—meaning that the model uses the noise and idiosyncrasies in the training data to improve predictions (Bullock et al., 2022). This can be difficult to detect, as the algorithm may still have high predictive performance in internal validation. For example, Rossi et al. (2018), who studied 26 football players with 23 injury events, compared a single decision tree, random forest, and logistic regression; the single decision tree, most prone to overfitting, had seemingly the best predictive performance, and logistic regression, the least prone to overfitting, had the worst predictive performance. The successful results in the study have been suspected to be ungeneralizable (Lövdal et al., 2020; Theron, 2020). To reduce risk of overfitting, CARTs require higher sample sizes than regression (van der Ploeg et al., 2014).

Finally, CARTs are sensitive to cases where the event (injury) is rare compared to non-events (no injury), a situation known as class imbalance (Majumdar et al., 2022). For instance, in the 74 competitive runners studied in Lövdal et al. (2020), had 41 183 non-events and 583 injury events, a distribution not uncommon in sports injury data (Bahr, 2016). The best CART classifier would classify all 41 766 observations as a non-event, and it would have classified 99% events correctly—thereby never predicting an event. In Lövdal et al. (2020) and in four football studies reviewed in Majumdar et al. (2022), sampling approaches were used to balance the event to non-event ratio. However, using such approaches, the distribution of injuries to non-injuries no longer represent the true distribution in the target population, and the corresponding classification is therefore less likely to predict correctly in external validation and in real-life applications (Goorbergh et al., 2022).

Neural networks are a branch of machine learning that does not require the dichotomization of continuous variables. Unlike classification methods, the output can be continuous, and can also be the probability of the event (injury) occurring, which potentially, a coach or clinician can act upon. To my knowledge, no study has so far attempted neural networks in the training load and injury risk field. One avenue where neural networks may be advantageous over regression modelling, is to search for time periods relative to previous time periods that can predict injury risk, which may improve our knowledge of whether relative training load effects are present further in the past. Whether, say, training load 3 to 6 days previously relative to 7 to 10 days previously predicts injury. While specific time period cut-offs are likely to be arbitrary and data-set specific, such a data-driven approach can provide hypothesis-generating approximations of time periods—an indication of day-to-day differences, 3-day relative differences, or 10 day relative differences, that could later be considered in studies of causal inference. Although, neural networks are at increased risk of overfitting (even more so than CARTs) and require considerable amounts of data, having steeper requirements than both logistic regression and classification methods (van der Ploeg et al., 2014). In a situation with big data (billions of rows) neural networks may provide the opportunity to study more complex relationships between variables than is feasible using traditional regression.

Given the sample sizes reported in *Paper I*, I argue that the training load and injury risk field must first improve data collection procedures to warrant the use of neural networks. In *Paper II*, *Paper III* and *Paper IV*, we demonstrate that regression analyses with proper model specification can model non-linearity, time-dependent effects and interactions, respectively. Logistic regression and time-to-event regression are probability-based methods which can handle imbalanced outcomes, are data efficient, and provide predicted probabilities which can inform practitioners. Machine learning approaches (herein CART and neural networks) should be used if they meet the aims and assumptions of the study, and not solely for their capability to handle non-linearity or other complexities.

4.7 Methodological considerations

4.7.1 Methodological considerations in simulations

In *Paper I–III*, statistical methods were compared in estimating the relationship between training load and injury under various circumstances. To feasibly report and interpret results, we limited the number of methods compared in each study by selecting from a variety of available approaches. We prioritized methods used frequently in the training load injury risk field, methods that were recommended but suspected to underperform, as well as currently proposed methods. Consequently, there was little space for comparing less known methods that have shown potential in other fields of research. For instance, in *Paper I*, we only included one form of multiple imputation, predicted mean matching, though other varieties, such as hot deck imputation or random forest imputation (C. S. Wang, Tyrel et al., 2020), may also be suitable in a multiple imputation framework. In *Paper II*, it may have been fruitful to include logarithmic transformation and machine learning classification for handling non-linearity (Bittencourt et al., 2016; Xiao et al., 2011). This reduces the novelty of the findings in *Paper I–II*.

We chose a varying degree of realism in the three simulation studies. In *Paper I*, we kept the data as-is and simulated missing data—meaning that all correlations, measurement error and distributions of variables were retained. We could therefore answer whether these variables were sufficient for the performance of PMM and/or regression imputation in a real study. In *Paper II* we added correlations between sampled sRPE observations to simulate longitudinal data, added noise to simulate measurement error, and simulated two sample sizes. Here, we had apriori hypotheses of method performance: the question was not whether the methods could detect the non-linear or linear shapes, but which was most optimal in a training load and injury scenario. In contrast, we did not simulate any correlations between measures or measurement error in *Paper III*, since the abilities of the methods to detect the relationships were largely unknown. If we had added noise/correlations, there would be no way of knowing whether methods did not detect a relationship because they cannot detect it at all, or whether it was due to noise and/or correlations. Consequently, the methods' performances under more realistic conditions is unclear.

The statistical performance measures varied between the simulations conducted in this dissertation. Firstly, RMSE, used in all simulation studies, does not describe the bias and precision of a method other than that relative to other methods. The visualizations in *Paper II* and *Paper III* were necessary additions to understand whether methods that were, based on RMSE, more accurate than other methods, also had high accuracy. The percent bias used in *Paper I* did not have this limitation. In *Paper II*, conducted chronologically before *Paper I*, we chose some of the same measures as in Carey et al. (2018) for reproducibility. In *Paper III*, we followed methodology and recommendations in Gasparrini (2014). In hindsight, we should have used the raw and percent bias in these papers as well.

Secondly, coverage had varying usefulness in the studies. In *Paper II*, we initially considered using coverage of 95% confidence intervals as to gauge certainty and bias. This rewarded uncertain methods: even though there was a large discrepancy between prediction and the simulated observation, the confidence intervals could still overlap due to high uncertainty. Vice versa, it punished more certain methods that may have been less biased than the uncertain methods, but their narrow confidence intervals did not necessarily overlap with the true observation. We therefore changed to 95% prediction intervals. Coverage of 95% confidence is potentially more useful when methods have similarly broad confidence intervals. However, coverage still has the disadvantage that it asks a yes/no question when a continuous answer may be more useful: If CIs do not overlap with the true relationship, by how much does the CI miss the mark? This may potentially be answered with an estimate of the area under the curve between the true relationship and the high or low confidence interval, depending on the direction of deviation. In our simulations, a combination of the percent bias, and average width of confidence intervals, was more descriptive of the performance we were aiming to determine than coverage metrics.

4.7.2 Methodological considerations in observed data analysis

All data used in this dissertation had limitations. The Norwegian football cohorts had few injury events, which likely made estimates highly uncertain and non-significant. The Norwegian elite youth handball data had poor timeliness and high amounts of missing data, possibly under missing not at random, which introduces selection bias. Finally, the Qatar Stars League cohort had no source of training load measure that describes the intensity of the activity, only the duration. This likely added noise to the model estimates.

4.8 Future directions

4.8.1 Future research in statistics for training load-injury research

Distributed lag non-linear models have shown potential as a method for determining the effect of training load on the risk of injury. However, little is known about how sensitive the method is to measurement error and limited sample size. Our results in *Paper I* and *Paper III* suggest that an accurate depiction of the relationship between training load and injury may require larger sample sizes than available in current literature, and DLNM may require even more to model detailed non-linear changes in the time-dependent effects. In addition, to use DLNM, the lag-period of effect must be defined before-hand. In our analysis of the handball population in *Paper III* and the football populations in *Paper IV*, we assumed that the previous four weeks of training load affected injury risk, and that any training load sustained before those four weeks had no effect. This is a strong assumption. Ideally, researchers can explore how far into the past training load affects injury risk before it is clinically irrelevant. Providing DLNM with a lag-period of a season, a year, or multiple years, will allow exploration of the decay of time-dependent effects. This is likely to require tremendous amounts of data. Defining how much data is needed to determine accurate effects of decay would be a beneficial avenue of research.

Collinearity is of concern for researchers in the space of training load monitoring and the field of training load and injury risk research (Weaving et al., 2020). Collinearity is a situation when highly correlated variables are included in a regression model, which inflate standard errors and therefore restrains the ability of the model to detect associations (Dormann et al., 2013). Coefficients can become volatile, and it can be difficult to separate the effect estimate from one variable and another (Dormann et al., 2013). Often, researchers gather multiple measures when describing the construct of training load. In football studies, for instance, this can be biomarkers, heart rate, athlete-reported intensity and multiple GPS measures (Miguel et al., 2021), and thus high between-measure correlations are expected. Collinearity may also present itself between time periods in past training load (Basagaña & Barrera-Gómez, 2021).

Coyne et al. (2022) suggested using data-driven approaches for variable selection, either AIC/R^2 for predictive ability, or dimension reduction methods such as principal components analysis (PCA). Williams, Trewartha, et al. (2017) and Weaving et al. (2020) recommended PCA and demonstrated how it could be used in rugby union. Thornton et al. (2017) used random forest, also in rugby union, to determine importance of different GPS variables in prediction of injury risk. It is unknown which approach is most feasible for causal inference and prediction, respectively, and how much data-driven approaches can be influenced by randomness. When independent variables in a model influence each other's estimates due to collinearity, it may cause data-driven selection of variables to become arbitrary (Harrell Jr, 2017). It is also unclear how correlated training load measures must be before they influence coefficient estimates. Applying a simulation may show how severe the issue of collinearity is in a training load approach and provide recommendations for how to handle collinearity between typical measures of training load.

4.8.2 Bridging the gap between research and practice

The statistical approaches promoted in this dissertation are targeted towards sport scientists, sport biostatisticians, and clinical researchers, to improve statistical methodology in *research*. This dissertation neither addressed implementing training load research to sports medicine and sports science practice (Finch, 2006), nor addressed statistical methodology in the field of quality improvement (Wheeler & Chambers, 2010). Research and quality improvement are different paradigms of analysis (Reinhardt & Ray, 2003).

In the quality improvement field, the current best practice (as dictated by research) is implemented to increase performance and mitigate injuries. Here, it is beneficial to determine how closely practice follows current guidelines, by monitoring so-called quality indicators—metrics that contribute to sporting success (Provost & Murray, 2011; Wheeler & Chambers, 2010). Unlike in research, where as much as data as possible is preferred to generalize findings, in quality improvement, narrowing down time series to the athlete level can be more informative than overviews of the whole, such as a football team (Provost & Murray, 2011; Ward et al., 2018). Because of the overarching differences between research and quality improvement, the statistical methodology considered best practice in research is not necessarily transferable to quality improvement in a real-world setting (Nijman et al., 2020; Sands et al., 2017).

Some studies have claimed that load monitoring—the process of surveying training load in real-time—can be used by coaches and athletes to both improve performance and reduce injury risk (Blanch & Gabbett, 2016; Hamlin et al., 2019), meaning, it can be used for continuous quality improvement. Such studies have so far only recommend monitoring variables *associated* with injury and/or performance. Without studies of causal inference, it is unclear whether these training load variables represent a construct that can be modified to improve practice, and even much less clear *how* these may be modified. Therefore, based on current knowledge in the training load and injury risk area, it is uncertain whether it is worth the limited time of practitioners in the real-world setting to collect and monitor the above-mentioned variables. The statistical methodology recommended in this dissertation has opened barriers to performing the research needed to understand how training load affects injury risk. Studies of causal inference in training load in injury risk are warranted.

So far, studies on training load monitoring have approached the subject through the lens of research, where thorough data collection and testing procedures are warranted (Haller et al., 2022). To my knowledge, no study has assessed how to monitor training load to achieve desired outcomes using the perspective of quality improvement. This is a low hanging fruit of research that may partially explain why so few injury prevention measures and strategies have been successfully implemented (O'Brien et al., 2017). Research has shown that athletes and coaches are positive to adopt injury prevention programs, but struggle to maintain compliance once it has been implemented (Harøy et al., 2019; McCall et al., 2016).

Research requires rigorous data collection with high demands for accuracy and precision (Reinhard & Ray, 2003). Quality improvement can make do with a handful of the most important metrics, ideally these are cost-efficient and easy to collect for daily, weekly, or monthly trend analyses (Provost & Murray, 2011). The data are collected and analyzed continuously, maybe in real-time, which may also have different statistical requirements than the finite data collection done for research (Nijman et al., 2020). In addition, the context of the target population must be considered (Finch, 2006), without necessarily aiming to generalize the findings. Use of quality improvement approaches may potentially improve compliance and adherence to injury prevention interventions.

In summary, to bridge the gap between research and sports medicine and sports science practice, studies are needed to 1) determine the causal relationships between training load and injury risk and 2) apply quality improvement approaches in sport settings.

5 Conclusion

This dissertation addresses key areas in how to study the relationship between training load and injury risk. Tangible recommendations can be made for these areas.

1) Missing data in continuous training load measures should be imputed with multiple imputation. If resources for performing multiple imputation are not available, complete case analysis is preferred over mean imputation, unless this would result in the deletion of a large proportion of injuries in the data.

2) Non-linearity should be explored with fractional polynomials or restricted cubic splines, where the choice of one over the other is a matter of preference and nuances in the data.

3) The cumulative effect of long-term training load on injury risk can be modelled with distributed lag non-linear models. If resources for performing DLNM are not available, the EWMA is, until further research proves otherwise, the second-best alternative.

4) Instead of using an acute:chronic workload ratio, relative training load can be assessed by modelling acute and chronic loads separately.

The use of recommendations presented in this dissertation will, hopefully, lead towards a consensus on how to analyze training load and injury risk, which would reduce multiple testing issues, improve interpretation, and finally, allow between-study comparisons and meta-analyses.

6 References

Ahmun, R., McCaig, S., Tallent, J., Williams, S., & Gabbett, T. (2019). Association of Daily Workload, Wellness, and Injury and Illness During Tours in International Cricketers. *International journal of sports physiology and performance*, *14*(3), 369-377. https://doi.org/10.1123/ijspp.2018-0315

Akenhead, R., & Nassis, G. P. (2016). Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *International journal of sports physiology and performance*, *11*(5), 587-593. https://doi.org/10.1123/ijspp.2015-0331

Albrecht, J., Biese, K. M., Bell, D. R., Schaefer, D. A., & Watson, A. M. (2020). Training Load and Injury Among Middle School–Aged Athletes. *Journal of athletic training*, *55*(9), 954-959. https://doi.org/10.4085/1062-6050-435-19

Alder, G. (2018). *diagrams.net*. In (Version 20.2.7) [Computer program]. JGraph Ltd. https://app.diagrams.net/

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, *33*(1), 178-196. https://doi.org/10.2307/1911894

Ambler, G., & Benner, A. (2015). *mfp: Multivariable Fractional Polynomials*. In (Version 1.5.2) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=mfp

Andrade, R., Wik, E. H., Rebelo-Marques, A., Blanch, P., Whiteley, R., Espregueira-Mendes, J., & Gabbett, T. J. (2020). 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*, *50*(9), 1613-1635. https://doi.org/10.1007/s40279-020-01308-6

Arazi, H., Asadi, A., Khalkhali, F., Boullosa, D., Hackney, A. C., Granacher, U., & Zouhal, H. (2020). Association Between the Acute to Chronic Workload Ratio and Injury Occurrence in Young Male Team Soccer Players: A Preliminary Study. *Frontiers in physiology*, *11*. https://doi.org/10.3389/fphys.2020.00608

Armstrong, B. (2006). Models for the relationship between ambient temperature and daily mortality. *Epidemiology*, *17*(6), 624-631. https://doi.org/10.1097/01.ede.0000239732.50999.8f

Auguié, B. (2019). *egg: Extensions for 'ggplot2': Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size*. In (Version 0.4.5) [R package]. https://CRAN.R-project.org/package=egg

Bache-Mathiesen, L. K. (2021a). *Missing data in training load: R code repository*. In (Version 1.0) [R code repository]. GitHub. https://github.com/lenakba/missing-data-in-training-load

Bache-Mathiesen, L. K. (2021b). *Modelling non-linearity in training load and injury research: R code repository*. In (Version 1.0) [R code repository]. GitHub. https://github.com/lenakba/load-injury-non-linearity-study

References

Bache-Mathiesen, L. K. (2021c). *Training load and injury studies data file (2021)* Version 1.0) GitHub. https://github.com/lenakba/missing-data-in-training-load/blob/main/studies_missing_reporting.xlsx

Bache-Mathiesen, L. K. (2022a). *Modelling the acute and chronic loads separately: R code repository*. In (Version 1.0) [R code repository]. https://github.com/lenakba/separating-acute-from-chronic

Bache-Mathiesen, L. K. (2022b). *Modelling the cumulative effect of training load on injury risk: R code repository*. In (Version 1.0) [R code repository]. https://github.com/lenakba/modelling-training-load

Bache-Mathiesen, L. K. (2022c). *Training load and injury studies data file (2022)* Version 2.0) GitHub. https://github.com/lenakba/Training-load-review-data/blob/main/tl_injury_studies.xlsx

Bahr, R. (2016). Why screening tests to predict injury do not work—and probably never will…: a critical review. *British journal of sports medicine*, *50*(13), 776-780. https://doi.org/10.1136/bjsports-2016-096256

Bahr, R., Clarsen, B., Derman, W., Dvorak, J., Emery, C. A., Finch, C. F., Hägglund, M., Junge, A., & Kemp, S. (2020). International Olympic Committee Consensus Statement: Methods for Recording and Reporting of Epidemiological Data on Injury and Illness in Sports 2020 (Including the STROBE Extension for Sports Injury and Illness Surveillance (STROBE-SIIS)). *Orthopaedic journal of sports medicine*, *8*(2). https://doi.org/10.1177/2325967120902908

Bahr, R., & Holme, I. (2003). Risk factors for sports injuries—a methodological approach. *British journal of sports medicine*, *37*(5), 384-392. https://doi.org/10.1136/bjsm.37.5.384

Bahr, R., & Krosshaug, T. (2005). Understanding injury mechanisms: a key component of preventing injuries in sport. *British journal of sports medicine*, *39*(6), 324-329. https://doi.org/10.1136/bjsm.2005.018341

Banister, E., Calvert, T., Savage, M., & Bach, T. (1975). A systems model of training for athletic performance. *Australian journal of sports medicine*, *7*(3), 57-61. https://doi.org/10.1109/TSMC.1976.5409179

Barnett, A. G., McElwee, P., Nathan, A., Burton, N. W., & Turrell, G. (2017). Identifying patterns of item missing survey data using latent groups: an observational study. *BMJ open*, *7*(10), e017284. https://doi.org/10.1136/bmjopen-2017-017284

Barzi, F., & Woodward, M. (2004). Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *American journal of epidemiology*, *160*(1), 34-45. https://doi.org/10.1093/aje/kwh175

Basagaña, X., & Barrera-Gómez, J. (2021). Reflection on modern methods: visualizing the effects of collinearity in distributed lag models. *International journal of epidemiology*, *51*(1), 334-344. https://doi.org/10.1093/ije/dyab179

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67(1)*, 1-48. https://doi.org/10.18637/jss.v067.i01

Bekker, S. (2019). Shuffle methodological deck chairs or abandon theoretical ship? The complexity turn in injury prevention. *Injury prevention*, *25*(2), 80-82. https://doi.org/10.1136/injuryprev-2018-042905

References

Benson, L. C., Owoeye, O. B., Räisänen, A. M., Stilling, C., Edwards, W. B., & Emery, C. A. (2021). Magnitude, Frequency, and Accumulation: Workload Among Injured and Uninjured Youth Basketball Players [Original Research]. *Frontiers in sports and active living*, *3*. https://doi.org/10.3389/fspor.2021.607205

Benson, L. C., Räisänen, A. M., Volkova, V. G., Pasanen, K., & Emery, C. A. (2020). Workload a-WEAR-ness: monitoring workload in team sports with wearable technology. a scoping review. *Journal of Orthopaedic & Sports Physical Therapy*, *50*(10), 549-563. https://doi.org/10.2519/jospt.2020.9753

Benson, L. C., Stilling, C., Owoeye, O. B., & Emery, C. A. (2021). Evaluating Methods for Imputing Missing Data from Longitudinal Monitoring of Athlete Workload. *Journal of Sports Science and Medicine*, *20*(2), 187-195. https://doi.org/10.52082/jssm.2021.188

Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L., & Armstrong, B. (2013). Time series regression studies in environmental epidemiology. *International journal of epidemiology*, *42*(4), 1187-1195. https://doi.org/10.1093/ije/dyt092

Binder, H., Sauerbrei, W., & Royston, P. (2013). Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. *Statistics in medicine*, *32*(13), 2262-2277. https://doi.org/10.1002/sim.5639

Bittencourt, N., Meeuwisse, W., Mendonça, L., Nettel-Aguirre, A., Ocarino, J., & Fonseca, S. (2016). Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *British journal of sports medicine*, *50*(21), 1309-1314. https://doi.org/10.1136/bjsports-2015-095850

Bjørndal, C. T., Bache-Mathiesen, L. K., Gjesdal, S., Moseid, C. H., Myklebust, G., & Luteberget, L. S. (2021). An Examination of Training Load, Match Activities, and Health Problems in Norwegian Youth Elite Handball Players Over One Competitive Season [Original Research]. *Frontiers in sports and active living*, *3*(36). https://doi.org/10.3389/fspor.2021.635103

Blanch, P., & Gabbett, T. J. (2016). 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*, *50*(8), 471-475. https://doi.org/10.1136/bjsports-2015-095445

Borg, D. N., Nguyen, R., & Tierney, N. J. (2022). Missing Data: Current Practice in Football Research and Recommendations for Improvement. *Science and Medicine in Football*, *6*(2), 262-267. https://doi.org/10.1080/24733938.2021.1922739

Borg, G., Hassmén, P., & Lagerström, M. (1987). Perceived exertion related to heart rate and blood lactate during arm and leg exercise. *European journal of applied physiology and occupational physiology*, *56*(6), 679-685. https://doi.org/10.1007/BF00424810

Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M. C., Gabbett, T. J., Coutts, A. J., Burgess, D. J., & Gregson, W. (2017). Monitoring athlete training loads: consensus statement. *International journal of sports physiology and performance*, *12*(s2), S2-161-S162-170. https://doi.org/10.1123/IJSPP.2017-0208

Bowen, L., Gross, A. S., Gimpel, M., Bruce-Low, S., & Li, F.-X. (2020). Spikes in acute: chronic workload ratio (ACWR) associated with a 5–7 times greater injury rate in English Premier

League football players: a comprehensive 3-year study. *British journal of sports medicine*, *54*, 731-738. https://doi.org/10.1136/bjsports-2018-099422

Boyd, L. J., Ball, K., & Aughey, R. J. (2011). The reliability of MinimaxX accelerometers for measuring physical activity in Australian football. *International journal of sports physiology and performance*, *6*(3), 311-321. https://doi.org/10.1123/ijspp.6.3.31121911857

Brankovic, M., Kardys, I., Steyerberg, E. W., Lemeshow, S., Markovic, M., Rizopoulos, D., & Boersma, E. (2019). Understanding of interaction (subgroup) analysis in clinical trials. *European Journal of Clinical Investigation*, *49*(8), e13145. https://doi.org/https://doi.org/10.1111/eci.13145

Brink, M. S., Nederhof, E., Visscher, C., Schmikli, S. L., & Lemmink, K. A. (2010). Monitoring load, recovery, and performance in young elite soccer players. *The Journal of Strength & Conditioning Research*, *24*(3), 597-603. https://doi.org/10.1519/JSC.0b013e3181c4d38b

Bullock, G. S., Hughes, T., Arundale, A. H., Ward, P., Collins, G. S., & Kluzek, S. (2022). Black Box Prediction Methods in Sports Medicine Deserve a Red Card for Reckless Practice: A Change of Tactics is Needed to Advance Athlete Care. *Sports Medicine*, *52*(8), 1729-1735. https://doi.org/10.1007/s40279-022-01655-6

Bullock, G. S., Hughes, T., Sergeant, J. C., Callaghan, M. J., Riley, R., & Collins, G. (2021). Methods matter: clinical prediction models will benefit sports medicine practice, but only if they are properly developed and validated. *British journal of sports medicine*, *55*(23), 1319-1321. https://doi.org/10.1136/bjsports-2021-104329

Buuren, S. v. G.-O., Karin. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1-67. https://doi.org/10.18637/jss.v045.i03

Cameron, K. L. (2010). COMMENTARY: Time for a Paradigm Shift in Conceptualizing Risk Factors in Sports Injury Research. *Journal of athletic training*, *45*(1), 58-60. https://doi.org/10.4085/1062-6050-45.1.58

Carbone, L., Sampietro, M., Cicognini, A., García-Sillero, M., & Vargas-Molina, S. (2022). Is the Relationship between Acute and Chronic Workload a Valid Predictive Injury Tool? A Bayesian Analysis. *Journal of Clinical Medicine*, *11*(19), 5945. https://doi.org/10.3390/jcm11195945

Carey, D. L., Blanch, P., Ong, K.-L., Crossley, K. M., Crow, J., & Morris, M. E. (2017). Training loads and injury risk in Australian football—differing acute: chronic workload ratios influence match injury risk. *British journal of sports medicine*, *51*(16), 1215-1220. https://doi.org/10.1136/bjsports-2016-096309

Carey, D. L., Crossley, K. M., Whiteley, R., Mosler, A., Ong, K.-L., Crow, J., & Morris, M. E. (2018). Modeling Training Loads and Injuries: The Dangers of Discretization. *Medicine and science in sports and exercise*, *50*(11), 2267-2276. https://doi.org/10.1249/MSS.0000000000001685

Casals, M., & Finch, C. F. (2017). Sports Biostatistician: a critical member of all sports science and medicine teams for injury prevention. *Injury prevention*, *23*(6), 423-427. https://doi.org/10.1136/injuryprev-2016-042211

Chhabra, G., Vashisht, V., & Ranjan, J. (2017). A comparison of multiple imputation methods for data with missing values. *Indian Journal of Science and Technology*, *10*(19), 1-7. https://doi.org/10.17485/ijst/2017/v10i19/110646

References

Claudino, J. G., de Oliveira Capanema, D., de Souza, T. V., Serrão, J. C., Pereira, A. C. M., & Nassis, G. P. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Medicine-Open*, *5*(1), 28. https://doi.org/10.1186/s40798-019-0202-3

Colby, M. J., Dawson, B., Peeling, P., Heasman, J., Rogalski, B., Drew, M. K., Stares, J., Zouhal, H., & Lester, L. (2017). Multivariate modelling of subjective and objective monitoring data improve the detection of non-contact injury risk in elite Australian footballers. *Journal of science and medicine in sport*, *20*(12), 1068-1074. https://doi.org/10.1016/j.jsams.2017.05.010

Collins, G. S., Ogundimu, E. O., Cook, J. A., Manach, Y. L., & Altman, D. G. (2016). Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Statistics in medicine*, *35*(23), 4124-4135. https://doi.org/10.1002/sim.6986

Coyne, J. O. C., Coutts, A. J., Newton, R. U., & Haff, G. G. (2022). The Current State of Subjective Training Load Monitoring: Follow-Up and Future Directions. *Sports Medicine - Open*, *8*(1), 53. https://doi.org/10.1186/s40798-022-00433-y

Coyne, J. O. C., Newton, R. U., & Haff, G. G. (2021). Relationships Between Internal Training Load in a Taper With Elite Weightlifting Performance Calculated Using Different Moving Average Methods. *International journal of sports physiology and performance*, *16*(3), 342-352. https://doi.org/10.1123/ijspp.2020-0002

Cross, M. J., Williams, S., Trewartha, G., Kemp, S. P., & Stokes, K. A. (2016). The influence of in-season training loads on injury risk in professional rugby union. *International journal of sports physiology and performance*, *11*(3), 350-355. https://doi.org/10.1123/ijspp.2015-0187

Curran-Everett, D. (2013). Explorations in statistics: the analysis of ratios and normalized data. *Advances in physiology education*, *37*(3), 213-219. https://doi.org/10.1152/advan.00053.2013

Curran-Everett, D., & Williams, C. L. (2015). Explorations in statistics: the analysis of change. *Advances in physiology education*, *39*(2), 49-54. https://doi.org/10.1152/advan.00018.2015

Dalen-Lorentsen, T. (2021). *Training load and health problems in football: More complex than we first thought?* [Doctoral thesis, Norwegian School of Sport Sciences]. Oslo. https://hdl.handle.net/11250/2834827

Dalen-Lorentsen, T., Andersen, T. E., Bjørneboe, J., Vagle, M., Martin, K. N., Kleppen, M., Fagerland, M. W., & Clarsen, B. (2021). A Cherry, Ripe for Picking: The Relationship Between the Acute-Chronic Workload Ratio and Health Problems. *Journal of Orthopaedic & Sports Physical Therapy*, *51*(4), 162-173. https://doi.org/10.2519/jospt.2021.9893

Dalen-Lorentsen, T., Bjørneboe, J., Clarsen, B., Vagle, M., Fagerland, M. W., & Andersen, T. E. (2021). Does load management using the acute:chronic workload ratio prevent health problems? A cluster randomised trial of 482 elite youth footballers of both sexes. *British journal of sports medicine*, *55*(2), 108-114. https://doi.org/10.1136/bjsports-2020-103003

Datatilsynet. (2017). *The anonymisation of personal data*. Norwegian Data Protection Agency. Retrieved 2022-11-07 from https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/anonymisation/

de Leeuw, A.-W., van der Zwaard, S., van Baar, R., & Knobbe, A. (2022). Personalized Machine Learning Approach to Injury Monitoring in Elite Volleyball Players. *European Journal of Sport Science*, *22*(4), 511-520. https://doi.org/10.1080/17461391.2021.1887369

References

Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: a simulation assessment. *Journal of Statistical Computation and Simulation*, *78*(1), 69-84. https://doi.org/10.1080/10629360600903866

Dennis, R., Farhart, R., Goumas, C., & Orchard, J. (2003). Bowling workload and the risk of injury in elite cricket fast bowlers. *Journal of science and medicine in sport*, *6*(3), 359-367. https://doi.org/10.1016/S1440-2440(03)80031-2

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, *45*(2), 265-282. https://doi.org/10.1111/j.2044-8317.1992.tb00992.x

Díaz-Ordaz, K., Kenward, M. G., Cohen, A., Coleman, C. L., & Eldridge, S. (2014). Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines. *Clinical Trials*, *11*(5), 590-600. https://doi.org/10.1177/1740774514537136

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27-46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Drew, M. K., Blanch, P., Purdam, C., & Gabbett, T. J. (2017). 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*, *51*(7), 618-619. https://doi.org/10.1136/bjsports-2016-096609

Eckard, T. G., Padua, D. A., Hearn, D. W., Pexa, B. S., & Frank, B. S. (2018). The relationship between training load and injury in athletes: a systematic review. *Sports Medicine*, *48*(8), 1929-1961. https://doi.org/10.1007/s40279-018-0951-z

Enright, K., Green, M., Hay, G., & Malone, J. J. (2020). Workload and Injury in Professional Soccer Players: Role of Injury Tissue Type and Injury Severity. *International journal of sports medicine*, *41*(02), 89-97. https://doi.org/10.1055/a-0997-6741

Esmaeili, A., Hopkins, W. G., Stewart, A. M., Elias, G. P., Lazarus, B. H., & Aughey, R. J. (2018). The individual and combined effects of multiple factors on the risk of soft tissue non-contact injuries in elite team sport athletes. *Frontiers in physiology*, *9*. https://doi.org/10.3389/fphys.2018.01280

Fanchini, M., Rampinini, E., Riggio, M., Coutts, A. J., Pecci, C., & McCall, A. (2018). Despite association, the acute: chronic work load ratio does not predict non-contact injury in elite footballers. *Science and Medicine in Football*, *2*(2), 108-114. https://doi.org/10.1080/24733938.2018.1429014

Finch, C. (2006). A new framework for research leading to sports injury prevention. *Journal of science and medicine in sport*, *9*(1), 3-9. https://doi.org/10.1016/j.jsams.2006.02.009

Finch, C. F., & Cook, J. (2014). Categorising sports injuries in epidemiological studies: the subsequent injury categorisation (SIC) model to address multiple, recurrent and exacerbation of injuries. *British journal of sports medicine*, *48*(17), 1276-1280. https://doi.org/10.1136/bjsports-2012-091729

References

Fonseca, S. T., Souza, T. R., Verhagen, E., Van Emmerik, R., Bittencourt, N. F., Mendonça, L. D., Andrade, A. G., Resende, R. A., & Ocarino, J. M. (2020). Sports injury forecasting and complexity: a synergetic approach. *Sports Medicine*, 1-14. https://doi.org/10.1007/s40279-020-01326-4

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A New Approach to Monitoring Exercise Training. *Journal of strength and conditioning research*, *15*(1), 109-115. https://europepmc.org/article/med/11708692

Frøslie, K. F., Røislien, J., Laake, P., Henriksen, T., Qvigstad, E., & Veierød, M. B. (2010). Categorisation of continuous exposure variables revisited. A response to the Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) Study. *BMC medical research methodology*, *10*(1), 103. https://doi.org/10.1186/1471-2288-10-103

Fuller, C. W., Ekstrand, J., Junge, A., Andersen, T. E., Bahr, R., Dvorak, J., Hägglund, M., McCrory, P., & Meeuwisse, W. H. (2006). Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Scandinavian journal of medicine & science in sports*, *16*(2), 83-92. https://doi.org/10.1111/j.1600-0838.2006.00528.x

Gabbett, T. J. (2016). The training—injury prevention paradox: should athletes be training smarter and harder? *British journal of sports medicine*, *50*(5), 273-280. https://doi.org/10.1136/bjsports-2015-095788

Gabbett, T. J. (2018). Debunking the myths about training load, injury and performance: empirical evidence, hot topics and recommendations for practitioners. *British journal of sports medicine*, *54*, 58-66. https://doi.org/10.1136/bjsports-2018-099784

Gabbett, T. J., Hulin, B. T., Blanch, P., & Whiteley, R. (2016). High training workloads alone do not cause sports injuries: how you get there is the real issue. *British journal of sports medicine*, *50*, 444-445. https://doi.org/10.1136/bjsports-2015-095567

Gamble, P. (2013). Reducing injury in elite sport-is simply restricting workloads really the answer? [Review]. *New Zealand Journal of Sports Medicine*, *40*, 36-38. https://sportsmedicine.co.nz/wp-content/uploads/2016/02/NZJSM-Vol-40-Issue-1.pdf#page=38

Gamble, P., Chia, L., & Allen, S. (2020). The illogic of being data-driven: reasserting control and restoring balance in our relationship with data and technology in football. *Science and Medicine in Football*, *4*(4), 338-341. https://doi.org/10.1080/24733938.2020.1854842

Gasparrini, A. (2011). Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software*, *43*(8), 1. https://doi.org/10.18637/jss.v043.i08

Gasparrini, A. (2014). Modeling exposure–lag–response associations with distributed lag non-linear models. *Statistics in medicine*, *33*(5), 881-899. https://doi.org/10.1002/sim.5963

Gasparrini, A. (2016). Modelling lagged associations in environmental time series data: a simulation study. *Epidemiology*, *27*(6), 835-842. https://doi.org/10.1097/EDE.0000000000000533

Gauthier, J., Wu, Q., & Gooley, T. (2020). Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. *Bone Marrow Transplant*, *55*, 675-680. https://doi.org/10.1038/s41409-019-0679-x

General Assembly of the World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama*, *310*(20), 2191-2194. https://doi.org/10.1001/jama.2013.281053

References

Goorbergh, R. v. d., Smeden, M. v., Timmerman, D., & Calster, B. v. (2022). The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. *Methodology (stat.ME), arXiv:2202.09101*. https://doi.org/10.48550/arXiv.2202.09101

Griffin, A., Kenny, I. C., Comyns, T. M., & Lyons, M. (2020). The Association Between the Acute:Chronic Workload Ratio and Injury and its Application in Team Sports: A Systematic Review. *Sports Medicine*, *50*(3), 561-580. https://doi.org/10.1007/s40279-019-01218-2

Haller, N., Blumkaitis, J. C., Strepp, T., Schmuttermair, A., Aglas, L., Simon, P., Neuberger, E., Kranzinger, C., Kranzinger, S., O'Brien, J., Ergoth, B., Raffetseder, S., Fail, C., Düring, M., & Stöggl, T. (2022). Comprehensive training load monitoring with biomarkers, performance testing, local positioning data, and questionnaires - first results from elite youth soccer. *Front Physiol*, *13*, 1000898. https://doi.org/10.3389/fphys.2022.1000898

Hamlin, M. J., Wilkes, D., Elliot, C. A., Lizamore, C. A., & Kathiravel, Y. (2019). Monitoring training loads and perceived stress in young elite university athletes. *Frontiers in physiology*, *10*, 34. https://doi.org/10.3389/fphys.2019.00034

Han, Q.-J., Wu, H.-L., Cai, C.-B., Xu, L., & Yu, R.-Q. (2008). An ensemble of Monte Carlo uninformative variable elimination for wavelength selection. *Analytica Chimica Acta*, *612*(2), 121-125. https://doi.org/10.1016/j.aca.2008.02.032

Harrell Jr, F. E. (2017). Regression modeling strategies. *BIOS*, *330*, 2018. https://link.springer.com/book/10.1007/978-3-319-19425-7

Harrell Jr, F. E. (2019). *rms: Regression Modeling Strategies*. In (Version 5.1-5) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=rms

Harøy, J., Wiger, E. G., Bahr, R., & Andersen, T. E. (2019). Implementation of the Adductor Strengthening Programme: Players primed for adoption but reluctant to maintain — A cross-sectional study. *Scandinavian journal of medicine & science in sports*, *29*(8), 1092-1100. https://doi.org/10.1111/sms.13444

Hecksteden, A., Kellner, R., & Donath, L. (2022). Dealing with small samples in football research. *Science and Medicine in Football*, *6*(3), 389-397. https://doi.org/10.1080/24733938.2021.1978106

Hernán, M. A. (2018). The C-word: scientific euphemisms do not improve causal inference from observational data. *American journal of public health*, *108*(5), 616-619. https://doi.org/10.2105/AJPH.2018.304337

Hildebrandt, C., Oberhoffer, R., Raschner, C., Müller, E., Fink, C., & Steidl-Müller, L. (2020). Training load characteristics and injury and illness risk identification in elite youth ski racing: A prospective study. *Journal of sport and health science*, *10*(2), 230-236. https://doi.org/10.1016/j.jshs.2020.03.009

Hocking, T. D. (2021). *directlabels: Direct Labels for Multicolor Plots*. In (Version 2021.1.13) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=directlabels

Hoffman, D. T., Dwyer, D. B., Bowe, S. J., Clifton, P., & Gastin, P. B. (2020). Is injury associated with team performance in elite Australian football? 20 years of player injury and team performance data that include measures of individual player value. *British journal of sports medicine*, *54*(8), 475-479. https://doi.org/10.1136/bjsports-2018-100029

References

Holländer, N., Sauerbrei, W., & Schumacher, M. (2004). Confidence intervals for the effect of a prognostic factor after selection of an 'optimal' cutpoint. *Statistics in medicine*, *23*(11), 1701-1713. https://doi.org/10.1002/sim.1611

Hulin, B. T., & Gabbett, T. J. (2019). Indeed association does not equal prediction: the never-ending search for the perfect acute: chronic workload ratio. *British journal of sports medicine*, *53*, 144-145. https://doi.org/10.1136/bjsports-2018-099448

Hulin, B. T., Gabbett, T. J., Blanch, P., Chapman, P., Bailey, D., & Orchard, J. W. (2014). Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *British journal of sports medicine*, *48*(8), 708-712. https://doi.org/10.1136/bjsports-2013-092524

Hulin, B. T., Gabbett, T. J., Lawson, D. W., Caputi, P., & Sampson, J. A. (2016). The acute: chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *British journal of sports medicine*, *50*(4), 231-236. https://doi.org/10.1136/bjsports-2015-094817

Hägglund, M., Waldén, M., Bahr, R., & Ekstrand, J. (2005). Methods for epidemiological study of injuries to professional football players: developing the UEFA model. *British journal of sports medicine*, *39*(6), 340-346. https://doi.org/10.1136/bjsm.2005.018267

Hägglund, M., Waldén, M., Magnusson, H., Kristenson, K., Bengtsson, H., & Ekstrand, J. (2013). Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. *British journal of sports medicine*, *47*(12), 738-742. https://doi.org/10.1136/bjsports-2013-092771

IBM. (2020, 2020-04-16). *Pairwise vs. Listwise deletion: What are they and when should I use them?* IBM Support. Retrieved 2022-11-07 from https://www.ibm.com/support/pages/pairwise-vs-listwise-deletion-what-are-they-and-when-should-i-use-them

Ide, B. N., Silvatti, A. P., & Mota, G. R. (2021). Comment on "Training Load and Injury: Causal Pathways and Future Directions". *Sports Medicine*, *51*, 2449–2450. https://doi.org/10.1007/s40279-021-01525-7

Impellizzeri, F., Woodcock, S., Coutts, A. J., Fanchini, M., McCall, A., & Vigotsky, A. (2020). Acute to random workload ratio is 'as' associated with injury as acute to actual chronic workload ratio: time to dismiss ACWR and its components. *SportRxiv*. https://doi.org/10.31236/osf.io/e8kt4

Impellizzeri, F., Woodcock, S., McCall, A., Ward, P., & Coutts, A. J. (2019). The acute-chronic workload ratio-injury figure and its 'sweet spot'are flawed. *SportRxiv*. https://doi.org/10.31236/osf.io/gs8yu

Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2019). Internal and external training load: 15 years on. *International journal of sports physiology and performance*, *14*(2), 270-273. https://doi.org/10.1123/ijspp.2018-0935

Impellizzeri, F. M., Shrier, I., McLaren, S., Coutts, A. J., McCall, A., Slattery, K., Jeffries, A., & Kalkhoven, J. (2022). Understanding training load as exposure and dose. *SportRχiv*. https://doi.org/10.51224/SRXIV.186

Impellizzeri, F. M., Tenan, M. S., Kempton, T., Novak, A., & Coutts, A. J. (2020). Acute: Chronic Workload Ratio: Conceptual Issues and Fundamental Pitfalls. *International journal of sports physiology and performance*, *15*(6), 907-913. https://doi.org/10.1123/ijspp.2019-0864

References

Impellizzeri, F. M., Ward, P., Coutts, A. J., Bornn, L., & McCall, A. (2020a). Training Load and Injury: Part 1—The Devil Is in the Detail—Challenges to Applying the Current Research in the Training Load and Injury Field. *Journal of Orthopaedic & Sports Physical Therapy*, *50*(10), 574-576. https://doi.org/10.2519/jospt.2020.9675

Impellizzeri, F. M., Ward, P., Coutts, A. J., Bornn, L., & McCall, A. (2020b). Training Load and Injury: Part 2—Questionable Research Practices Hijack the Truth and Mislead Well-Intentioned Clinicians. *Journal of Orthopaedic & Sports Physical Therapy*, *50*(10), 577-584. https://doi.org/10.2519/jospt.2020.9211

Impellizzeri, F. M., Woodcock, S., Coutts, A., Fanchini, M., McCall, A., & Vigotsky, A. (2021). What Role Do Chronic Workloads Play in the Acute to Chronic Workload Ratio? Time to Dismiss ACWR and Its Underlying Theory. *Sports Medicine*, *51*, 581–592. https://doi.org/10.1007/s40279-020-01378-6

Information Commissioner's Office (ICO). (2018). *Guide to the UK General Data Protection Regulation (UK GDPR)*. Retrieved 2022-11-07 from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology*, *17*(1), 162. https://doi.org/10.1186/s12874-017-0442-1

James, D. H., Kurt. (2020). *chron: Chronological Objects which Can Handle Dates and Times*. In (Version version 2.3-56) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=chron

Jamil, M., Phatak, A., Mehta, S., Beato, M., Memmert, D., & Connor, M. (2021). Using multiple machine learning algorithms to classify elite and sub-elite goalkeepers in professional men's football. *Scientific Reports*, *11*(1), 22703. https://doi.org/10.1038/s41598-021-01187-5

Janssen, K. J., Donders, A. R. T., Harrell Jr, F. E., Vergouwe, Y., Chen, Q., Grobbee, D. E., & Moons, K. G. (2010). Missing covariate data in medical research: to impute is better than to ignore. *Journal of clinical epidemiology*, *63*(7), 721-727. https://doi.org/10.1016/j.jclinepi.2009.12.008

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental psychology*, *45*(4), 1195–1199. https://doi.org/10.1037/a0015665

Johansson, F., Gabbett, T., Svedmark, P., & Skillgate, E. (2021). External Training Load and the Association With Back Pain in Competitive Adolescent Tennis Players: Results From the SMASH Cohort Study. *Sports health*, *14*(1), 111-118. https://doi.org/10.1177/19417381211051636

Johnston, R., Cahalan, R., Bonnett, L., Maguire, M., Nevill, A., Glasgow, P., O'Sullivan, K., & Comyns, T. (2019). Training load and baseline characteristics associated with new injury/pain within an endurance sporting population: a prospective study. *International journal of sports physiology and performance*, *14*(5), 590-597. https://doi.org/10.1123/ijspp.2018-0644

References

Johnstone, I. M., & Titterington, D. M. (2009). Statistical challenges of high-dimensional data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences*, *367*, 4237–4253. https://doi.org/10.1098/rsta.2009.0159

Kabacoff, R. (2011). *R in Action*. Manning publications Shelter Island, NY, USA. https://www.manning.com/books/r-in-action-second-edition

Kalkhoven, J. T., Watsford, M. L., Coutts, A. J., Edwards, W. B., & Impellizzeri, F. M. (2021). Training load and injury: causal pathways and future directions. *Sports Medicine*, *51*, 1137–1150. https://doi.org/10.1007/s40279-020-01413-6

Karahalios, A., Baglietto, L., Carlin, J. B., English, D. R., & Simpson, J. A. (2012). A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. *BMC medical research methodology*, *12*(1), 96. https://doi.org/10.1186/1471-2288-12-96

Keylock, L., Alway, P., Felton, P., McCaig, S., Brooke-Wavell, K., King, M., & Peirce, N. (2022). Lumbar bone stress injuries and risk factors in adolescent cricket fast bowlers. *Journal of sports sciences*, *40*(12), 1336-1342. https://doi.org/10.1080/02640414.2022.2080161

Knowles, J. E., & Frederick, C. (2019). *merTools: Tools for Analyzing Mixed Effect Regression Models.* In (Version 0.5.0) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=merTools

Lathlean, T. J., Gastin, P. B., Newstead, S. V., & Finch, C. F. (2019). Absolute and relative load and injury in elite junior Australian football players over 1 season. *International journal of sports physiology and performance*, *15*(4), 511-519. https://doi.org/10.1123/ijspp.2019-0100

Lathlean, T. J. H. (2017). *Training Loads, Player Wellness and Injury in Elite Junior Australian Football Players* [Doctoral thesis, Monash University].

Lathlean, T. J. H., Newstead, S. V., & Gastin, P. B. (2022). Elite Junior Australian Football Players With Impaired Wellness Are at Increased Injury Risk at High Loads. *Sports health*, *0*(0). https://doi.org/10.1177/19417381221087245

Lazarus, B. H., Stewart, A. M., White, K. M., Rowell, A. E., Esmaeili, A., Hopkins, W. G., & Aughey, R. J. (2017). Proposal of a global training load measure predicting match performance in an elite team sport. *Frontiers in physiology*, *8*, 930. https://doi.org/10.3389/fphys.2017.00930

Lich, K. H., Ginexi, E. M., Osgood, N. D., & Mabry, P. L. (2013). A Call to Address Complexity in Prevention Science Research. *Prevention Science*, *14*(3), 279-289. https://doi.org/10.1007/s11121-012-0285-2

Lolli, L., Bahr, R., Weston, M., Whiteley, R., Tabben, M., Bonanno, D., Gregson, W., Chamari, K., Di Salvo, V., & van Dyk, N. (2020). No association between perceived exertion and session duration with hamstring injury occurrence in professional football. *Scandinavian journal of medicine & science in sports*, *30*(3), 523-530. https://doi.org/10.1111/sms.13591

Lolli, L., Batterham, A. M., Hawkins, R., Kelly, D. M., Strudwick, A. J., Thorpe, R., Gregson, W., & Atkinson, G. (2019). Mathematical coupling causes spurious correlation within the conventional acute-to-chronic workload ratio calculations. *British journal of sports medicine*, *53*(15), 921-922. https://doi.org/10.1136/bjsports-2017-098110

Lolli, L., Batterham, A. M., Hawkins, R., Kelly, D. M., Strudwick, A. J., Thorpe, R. T., Gregson, W., & Atkinson, G. (2018). The acute-to-chronic workload ratio: an inaccurate scaling index for an unnecessary normalisation process? *British journal of sports medicine*, *53*, 1510-1512. https://doi.org/10.1136/bjsports-2017-098884

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Lüdecke, D. (2022). *sjPlot: Data Visualization for Statistics in Social Science*. In (Version 2.8.11) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=sjPlot

Lövdal, S., den Hartigh, R., & Azzopardi, G. (2020). Injury Prediction in Competitive Runners with Machine Learning. *International journal of sports physiology and performance*, *16*(10), 1522–1531. https://doi.org/10.1123/ijspp.2020-0518

Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of clinical epidemiology*, *110*, 63-73. https://doi.org/10.1016/j.jclinepi.2019.02.016

Majumdar, A., Bakirov, R., Hodges, D., Scott, S., & Rees, T. (2022). Machine Learning for Understanding and Predicting Injuries in Football. *Sports Medicine - Open*, *8*(1), 73. https://doi.org/10.1186/s40798-022-00465-4

Mallo, J., & Dellal, A. (2012). Injury risk in professional football players with special reference to the playing position and training periodization. *The Journal of sports medicine and physical fitness*, *52*(6), 631-638. https://europepmc.org/article/med/23187326

Malone, J. J., Di Michele, R., Morgans, R., Burgess, D., Morton, J. P., & Drust, B. (2015). Seasonal Training-Load Quantification in Elite English Premier League Soccer Players. *International journal of sports physiology and performance*, *10*(4), 489-497. https://doi.org/10.1123/ijspp.2014-0352

Malone, J. J., Lovell, R., Varley, M. C., & Coutts, A. J. (2017). Unpacking the black box: applications and considerations for using GPS devices in sport. *International journal of sports physiology and performance*, *12*(s2), S2-18-S12-26. https://doi.org/10.1123/ijspp.2016-0236

Malone, S., Collins, K., McROBERTS, A., & Doran, D. (2020). Understanding the association between external training load measures and injury risk in Elite Gaelic football. *The Journal of sports medicine and physical fitness*, *61*(2), 233-243. https://doi.org/10.23736/s0022-4707.20.11206-4

Malone, S., Hughes, B., Doran, D. A., Collins, K., & Gabbett, T. J. (2019). Can the workload–injury relationship be moderated by improved strength, speed and repeated-sprint qualities? *Journal of science and medicine in sport*, *22*(1), 29-34. https://doi.org/10.1016/j.jsams.2018.01.010

Malone, S., Owen, A., Mendes, B., Hughes, B., Collins, K., & Gabbett, T. J. (2018). High-speed running and sprinting as an injury risk factor in soccer: Can well-developed physical qualities reduce the risk? *Journal of science and medicine in sport*, *21*(3), 257-262. https://doi.org/10.1016/j.jsams.2017.05.016

Malone, S., Roe, M., Doran, D. A., Gabbett, T. J., & Collins, K. (2017). High chronic training loads and exposure to bouts of maximal velocity running reduce injury risk in elite Gaelic football. *Journal of science and medicine in sport*, *20*(3), 250-254. https://doi.org/10.1016/j.jsams.2016.08.005

Mandorino, M., Figueiredo, A. J., Cima, G., & Tessitore, A. (2022). Predictive Analytic Techniques to Identify Hidden Relationships between Training Load, Fatigue and Muscle

Strains in Young Soccer Players. *Sports*, *10*(1), 3. https://doi.org/10.3390/sports10010003

Mandorino, M., Figueiredo, A. J., Condello, G., & Tessitore, A. (2022). The influence of maturity on recovery and perceived exertion, and its relationship with illnesses and non-contact injuries in young soccer players. *Biology of sport*, *39*(4), 839-848. https://doi.org/10.5114/biolsport.2022.109953

Marshall, A., Altman, D. G., Royston, P., & Holder, R. L. (2010). Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC medical research methodology*, *10*(1), 7. https://doi.org/10.1186/1471-2288-10-7

Maupin, D., Schram, B., Canetti, E., & Orr, R. (2020). The Relationship Between Acute: Chronic Workload Ratios and Injury Risk in Sports: A Systematic Review. *Open Access Journal of Sports Medicine*, *11*, 51–75. https://doi.org/10.2147/OAJSM.S231405

McCall, A., Dupont, G., & Ekstrand, J. (2016). Injury prevention strategies, coach compliance and player adherence of 33 of the UEFA Elite Club Injury Study teams: a survey of teams' head medical officers. *British journal of sports medicine*, *50*(12), 725-730. https://doi.org/10.1136/bjsports-2015-095259

McCall, A., Dupont, G., & Ekstrand, J. (2018). Internal workload and non-contact injury: a one-season study of five teams from the UEFA Elite Club Injury Study. *British journal of sports medicine*, *52*(23), 1517-1522. https://doi.org/10.1136/bjsports-2017-098473

Meeuwisse, W. H. (1994). Assessing Causation in Sport Injury: A Multifactorial Model. *Clinical Journal of Sport Medicine*, *4*(3), 166-170. https://journals.lww.com/cjsportsmed/Fulltext/1994/07000/Assessing_Causation_in_Sport_Injury___A.4.aspx

Meeuwisse, W. H., Tyreman, H., Hagel, B., & Emery, C. (2007). A dynamic model of etiology in sport injury: the recursive nature of risk and causation. *Clinical Journal of Sport Medicine*, *17*(3), 215-219. https://doi.org/10.1097/JSM.0b013e3180592a48

Mehta, S. (2019). Relationship between workload and throwing injury in varsity baseball players. *Physical Therapy in Sport*, *40*, 66-70. https://doi.org/10.1016/j.ptsp.2019.08.001

Menaspà, P. (2017). Are rolling averages a good way to assess training load for injury prevention? *British journal of sports medicine*, *51*(7), 618-619. https://doi.org/10.1136/bjsports-2016-096131

Miguel, M., Oliveira, R., Loureiro, N., García-Rubio, J., & Ibáñez, S. J. (2021). Load Measures in Training/Match Monitoring in Soccer: A Systematic Review. *International Journal of Environmental Research and Public Health*, *18*(5), 2721. https://doi.org/10.3390/ijerph18052721

Moreno-Perez, V., Paredes, V., Pastor, D., Garrosa, F. N., Vielcazat, S. J., Del Coso, J., & Mendez-Villanueva, A. (2021). Under-exposure to official matches is associated with muscle injury incidence in professional footballers. *Biology of sport*, *38*(4), 563-571. https://doi.org/10.5114/biolsport.2021.100360

Moreno-Pérez, V., Prieto, J., Del Coso, J., Lidó-Micó, J. E., Fragoso, M., Penalva, F. J., Reid, M., & Pluim, B. M. (2021). Association of acute and chronic workloads with injury risk in high-performance junior tennis players. *European Journal of Sport Science*, *21*(8), 1215-1223. https://doi.org/10.1080/17461391.2020.1819435

References

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in medicine*, *38*(11), 2074-2102. https://doi.org/10.1002/sim.8086

Moussa, I., Leroy, A., Sauliere, G., Schipman, J., Toussaint, J.-F., & Sedeaud, A. (2019). Robust Exponential Decreasing Index (REDI): adaptive and robust method for computing cumulated workload. *BMJ Open Sport & Exercise Medicine*, *5*(1), e000573. https://doi.org/10.1136/bmjsem-2019-000573

Mujika, I., Padilla, S., Pyne, D., & Busso, T. (2004). Physiological Changes Associated with the Pre-Event Taper in Athletes. *Sports Medicine*, *34*(13), 891-927. https://doi.org/10.2165/00007256-200434130-00003

Murray, N. B., Gabbett, T. J., Townshend, A. D., & Blanch, P. (2017). Calculating acute: chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages. *British journal of sports medicine*, *51*(9), 749-754. https://doi.org/10.1136/bjsports-2016-097152

Mustapich, C. K. W., & Koehle, M. S. (2021). Effects of training load and non-training stress on injury risk in collegiate ice hockey players. *TRANSLATIONAL SPORTS MEDICINE*, *4*(6), 931-936. https://doi.org/10.1002/tsm2.297

Myklebust, G., & Bahr, R. (2005). Return to play guidelines after anterior cruciate ligament surgery. *British journal of sports medicine*, *39*(3), 127-131. https://doi.org/10.1136/bjsm.2004.010900

Myklebust, G., Holm, I., Mæhlum, S., Engebretsen, L., & Bahr, R. (2003). Clinical, Functional, and Radiologic Outcome in Team Handball Players 6 to 11 Years after Anterior Cruciate Ligament Injury:A Follow-up Study. *The American Journal of Sports Medicine*, *31*(6), 981-989. https://doi.org/10.1177/03635465030310063901

Nakaoka, G., Barboza, S. D., Verhagen, E., Van Mechelen, W., & Hespanhol, L. (2021). The association between the acute: chronic workload ratio and running-related injuries in Dutch runners: a prospective cohort study. *Sports Medicine*, *51*, 2437–2447. https://doi.org/10.1007/s40279-021-01483-0

Nielsen, R. O., Bertelsen, M. L., Møller, M., Hulme, A., Mansournia, M. A., Casals, M., & Parner, E. T. (2020). Methods matter: exploring the 'too much, too soon'theory, part 1: causal questions in sports injury research. *British journal of sports medicine*, *54*, 1119-1122. https://doi.org/10.1136/bjsports-2018-100245

Nielsen, R. O., Bertelsen, M. L., Ramskov, D., Møller, M., Hulme, A., Theisen, D., Finch, C. F., Fortington, L. V., Mansournia, M. A., & Parner, E. T. (2019). Time-to-event analysis for sports injury research part 1: time-varying exposures. *British journal of sports medicine*, *53*(1), 61-68. https://doi.org/10.1136/bjsports-2018-099408

Nielsen, R. O., Shrier, I., Casals, M., Nettel-Aguirre, A., Møller, M., Bolling, C., Bittencourt, N. F., Clarsen, B., Wedderkopp, N., & Soligard, T. (2020). Statement on Methods in Sport Injury Research From the First METHODS MATTER Meeting, Copenhagen, 2019. *Journal of Orthopaedic & Sports Physical Therapy*, *50*(5), 226-233. https://doi.org/10.1136/bjsports-2019-101323

Nielsen, R. O., Simonsen, N. S., Casals, M., Stamatakis, E., & Mansournia, M. A. (2020). Methods matter and the 'too much, too soon'theory (part 2): what is the goal of your sports injury research? Are you describing, predicting or drawing a causal inference?

British journal of sports medicine, *54*, 1307-1309. https://doi.org/10.1136/bjsports-2020-102144

Nijman, S. W. J., Hoogland, J., Groenhof, T. K. J., Brandjes, M., Jacobs, J. J. L., Bots, M. L., Asselbergs, F. W., Moons, K. G. M., & Debray, T. P. A. (2020). Real-time imputation of missing predictor values in clinical practice. *European Heart Journal - Digital Health*, *2*(1), 154-164. https://doi.org/10.1093/ehjdh/ztaa016

Nunes, A. C. L., Mendes, L. A., Mota, L. d. A., Lima, P. O. d. P., & Almeida, G. P. L. (2022). Training Load, Pain Intensity, and Functioning Can Explain Injuries in Dancers: A Classification and Regression Tree (CART) Analysis. *Medical Problems of Performing Artists*, *37*(2), 73-77. https://doi.org/10.21091/mppa.2022.2012

O'Brien, J., Young, W., & Finch, C. F. (2017). The use and modification of injury prevention exercises by professional youth soccer teams. *Scandinavian journal of medicine & science in sports*, *27*(11), 1337-1346. https://doi.org/10.1111/sms.12756

O'Kelly, M., Anisimov, V., Campbell, C., & Hamilton, S. (2017). Proposed best practice for projects that involve modelling and simulation. *Pharmaceutical statistics*, *16*(2), 107-113. https://doi.org/10.1002/pst.1789

Okholm Kryger, K., Dor, F., Guillaume, M., Haida, A., Noirez, P., Montalvan, B., & Toussaint, J.-F. (2015). Medical reasons behind player departures from male and female professional tennis competitions. *The American Journal of Sports Medicine*, *43*(1), 34-40. https://doi.org/10.1177/0363546514552996

Orchard, J. W., Meeuwisse, W., Derman, W., Hägglund, M., Soligard, T., Schwellnus, M., & Bahr, R. (2020). Sport Medicine Diagnostic Coding System (SMDCS) and the Orchard Sports Injury and Illness Classification System (OSIICS): revised 2020 consensus versions. *British journal of sports medicine*, *54*(7), 397-401. https://doi.org/10.1136/bjsports-2019-101921

Owen, A. L., Forsyth, J. J., Wong, D. P., Dellal, A., Connelly, S. P., & Chamari, K. (2015). Heart rate–based training intensity and its impact on injury incidence among elite-level professional soccer players. *The Journal of Strength & Conditioning Research*, *29*(6), 1705-1712. https://doi.org/10.1519/JSC.0000000000000810

Peng, R., & McDermott, A. (2022). *tsModel: Time Series Modeling for Air Pollution and Health*. In (Version 0.6-1) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=tsModel

Pepe, M. S., Janes, H., Longton, G., Leisenring, W., & Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American journal of epidemiology*, *159*(9), 882-890. https://doi.org/10.1093/aje/kwh101

Powney, M., Williamson, P., Kirkham, J., & Kolamunnage-Dona, R. (2014). A review of the handling of missing longitudinal outcome data in clinical trials. *Trials*, *15*(1), 237. https://doi.org/10.1186/1745-6215-15-237

Privacy EU. (2018). *General Data Protection Agency Article 13*. intersoft consulting. Retrieved 2022-11-07 from https://gdpr-info.eu/art-13-gdpr/

Provost, L. P., & Murray, S. K. (2011). *The health care data guide - Learning from data for improvement*. John Wiley Sons Inc. https://www.wiley.com/en-us/The+Health+Care+Data+Guide%3A+Learning+from+Data+for+Improvement-p-9780470902585

References

Pustejovsky, J. (2021). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections*. In (Version 0.5.3) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=clubSandwich

R Core Team. (2021). *R: A language and environment for statistical computing*. In [Computer program]. Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rago, V., Brito, J., Figueiredo, P., Costa, J., Barreira, D., Krustrup, P., & Rebelo, A. (2020). Methods to collect and interpret external training load using microtechnology incorporating GPS in professional football: a systematic review. *Research in Sports Medicine*, *28*(3), 437-458. https://doi.org/10.1080/15438627.2019.1686703

Ramskov, D., Rasmussen, S., Sørensen, H., Parner, E. T., Lind, M., & Nielsen, R. (2021). Interactions between running volume and running pace on injury occurrence in recreational runners: A secondary analysis. *Journal of athletic training*, *57*(6), 557–563. https://doi.org/10.4085/1062-6050-0165.21

Reinhardt, A. C., & Ray, L. N. (2003). Differentiating quality improvement from research. *Applied Nursing Research*, *16*(1), 2-8. https://doi.org/10.1053/apnr.2003.59000

Renfree, A., Casado, A., & McLaren, S. (2021). Re-thinking athlete training loads: would you rather have one big rock or lots of little rocks dropped on your foot? *Research in Sports Medicine*, *30*(5), 573-576. https://doi.org/10.1080/15438627.2021.1906672

Rhon, D. I., Teyhen, D. S., Collins, G. S., & Bullock, G. S. (2022). Predictive models for musculoskeletal injury risk: why statistical approach makes all the difference. *BMJ Open Sport & Exercise Medicine*, *8*(4), e001388. https://doi.org/10.1136/bmjsem-2022-001388

Richardson, D. B. (2009). Latency models for analyses of protracted exposures. *Epidemiology*, *20*(3), 395-399. https://doi.org/10.1097/EDE.0b013e318194646d

Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statistics in medicine*, *38*(7), 1276-1296. https://doi.org/10.1002/sim.7992

Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernàndez, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *Plos One*, *13*(7), e0201264. https://doi.org/10.1371/journal.pone.0201264

Royston, P., & Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *43*(3), 429-453. https://doi.org/10.2307/2986270

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*, *25*(1), 127-141. https://doi.org/10.1002/sim.2331

Ruddy, J. D., Cormack, S. J., Whiteley, R., Williams, M. D., Timmins, R. G., & Opar, D. A. (2019). Modeling the risk of team sport injuries: a narrative review of different statistical approaches. *Frontiers in physiology*, *10*, 829. https://doi.org/10.3389/fphys.2019.00829

Ryan, M. R., Napier, C., Greenwood, D., & Paquette, M. R. (2021). Comparison of different measures to monitor week-to-week changes in training load in high school runners. *International Journal of Sports Science & Coaching*, *16*(2), 370-379. https://doi.org/10.1177/1747954120970305

References

Rønneberg, K. L. (2020). *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* [Master thesis, Norwegian School of Sport Sciences]. https://hdl.handle.net/11250/2660778

Sainani, K. L., Borg, D. N., Caldwell, A. R., Butson, M. L., Tenan, M. S., Vickers, A. J., Vigotsky, A. D., Warmenhoven, J., Nguyen, R., Lohse, K. R., Knight, E. J., & Bargary, N. (2021). Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. *British journal of sports medicine*, *55*(2), 118-122. https://doi.org/10.1136/bjsports-2020-102607

Sampson, J. A., Fullagar, H., & Murray, A. (2017). Evidence is needed to determine if there is a better way to determine the acute: chronic workload. *British journal of sports medicine*, *51*(7), 621-622. https://doi.org/10.1136/bjsports-2016-097085

Sampson, J. A., Murray, A., Williams, S., Halseth, T., Hanisch, J., Golden, G., & Fullagar, H. (2018). Injury risk-workload associations in NCAA American college football. *Journal of science and medicine in sport*, *21*(12), 1215-1220. https://doi.org/10.1016/j.jsams.2018.05.019

Sampson, J. A., Murray, A., Williams, S., Sullivan, A., & Fullagar, H. H. (2019). Subjective Wellness, Acute: Chronic Workloads, and Injury Risk in College Football. *Journal of strength and conditioning research*, *33*(12), 3367-3373. https://doi.org/10.1519/JSC.0000000000003000

Sandnes, O. T. (2021). *Everything about football.* TV2. Retrieved 2022-11-07 from www.altomfotball.no

Sands, W. A., Kavanaugh, A. A., Murray, S. R., McNeal, J. R., & Jemni, M. (2017). Modern Techniques and Technologies Applied to Training and Performance Monitoring. *International journal of sports physiology and performance*, *12*(s2), S2-63-S62-72. https://doi.org/10.1123/ijspp.2016-0405

Saw, R., Dennis, R. J., Bentley, D., & Farhart, P. (2011). Throwing workload and injury risk in elite cricketers. *British journal of sports medicine*, *45*(10), 805-808. https://doi.org/10.1136/bjsm.2009.061309

Sedeaud, A., De Larochelambert, Q., Moussa, I., Brasse, D., Berrou, J.-M., Duncombe, S., Antero, J., Orhant, E., Carling, C., & Toussaint, J.-F. (2020). 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*, *11*, 1034. https://doi.org/10.3389/fphys.2020.01034

Shaw, J. W., Mattiussi, A. M., Brown, D. D., Williams, S., Kelly, S., Springham, M., Pedlar, C. R., & Tallent, J. (2021). Dance Exposure, Individual Characteristics, and Injury Risk over Five Seasons in a Professional Ballet Company. *Medicine & Science in Sports & Exercise*, *53*(11), 2290-2297. https://doi.org/10.1249/mss.0000000000002725

Shmueli, G. (2010). To explain or to predict? *Statistical science*, *25*(3), 289-310. https://doi.org/10.1214/10-STS330

Shrier, I., & Platt, R. W. (2008). Reducing bias through directed acyclic graphs. *BMC medical research methodology*, *8*(1), 70. https://doi.org/10.1186/1471-2288-8-70

Shrier, I., Wang, C., Stokes, T., Trejovargas, J., Stovitz, S. D., & Steele, R. J. (2022). Causal effects, workload and injury risk: The importance of specifying the research question.

Journal of science and medicine in sport, *25*(7), 574-578. https://doi.org/10.1016/j.jsams.2022.03.018

Shrier, I., Wang, C., Stokes, T., Vargas, J. T., & Steele, R. (2021). Important Nuances for Non-Linear Modeling [Commentary]. *BMJ Open Sport & Exercise Medicine*. https://bmjopensem.bmj.com/content/7/3/e001119.responses#important-nuances-for-non-linear-modeling

Siddique, J., Brown, C. H., Hedeker, D., Duan, N., Gibbons, R. D., Miranda, J., & Lavori, P. W. (2008). Missing Data in Longitudinal Trials – Part B, Analytic Issues. *Psychiatric Annals*, *38*(12). https://doi.org/10.3928/00485713-20081201-09

Sidey-Gibbons, J. A., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: a practical introduction. *BMC medical research methodology*, *19*(1), 64. https://doi.org/10.1186/s12874-019-0681-4

Skazalski, C., Whiteley, R., Hansen, C., & Bahr, R. (2018). A valid and reliable method to measure jump-specific training and competition load in elite volleyball players. *Scandinavian journal of medicine & science in sports*, *28*(5), 1578-1585. https://doi.org/10.1111/sms.13052

Sniffen, K., Noel-London, K., Schaeffer, M., & Owoeye, O. (2022). Is Cumulative Load Associated with Injuries in Youth Team Sport? A Systematic Review. *Sports Medicine - Open*, *8*(1), 117. https://doi.org/10.1186/s40798-022-00516-w

Soligard, T., Schwellnus, M., Alonso, J.-M., Bahr, R., Clarsen, B., Dijkstra, H. P., Gabbett, T., Gleeson, M., Hägglund, M., & Hutchinson, M. R. (2016). How much is too much?(Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *British journal of sports medicine*, *50*(17), 1030-1041. https://doi.org/10.1136/bjsports-2016-096581

Spanos, A., Harrell, F. E., Jr, & Durack, D. T. (1989). Differential Diagnosis of Acute Meningitis: An Analysis of the Predictive Value of Initial Observations. *Jama*, *262*(19), 2700-2707. https://doi.org/10.1001/jama.1989.03430190084036

Stares, J., Dawson, B., Peeling, P., Heasman, J., Rogalski, B., Drew, M., Colby, M., Dupont, G., & Lester, L. (2018). Identifying high risk loading conditions for in-season injury in elite Australian football players. *Journal of science and medicine in sport*, *21*(1), 46-51. https://doi.org/10.1016/j.jsams.2017.05.012

Staunton, C. A., Abt, G., Weaving, D., & Wundersitz, D. W. T. (2021). Misuse of the term 'load' in sport and exercise science. *Journal of science and medicine in sport*. https://doi.org/10.1016/j.jsams.2021.08.013

Stern, B. D., Hegedus, E. J., & Lai, Y.-C. (2020). Injury prediction as a non-linear system. *Physical Therapy in Sport*, *41*, 43-48. https://doi.org/10.1016/j.ptsp.2019.10.010

Stern, B. D., Hegedus, E. J., & Lai, Y.-C. (2021). State dependence: Does a prior injury predict a future injury? *Physical Therapy in Sport*, *49*, 8-14. https://doi.org/10.1016/j.ptsp.2021.01.008

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, *338*, b2393. https://doi.org/10.1136/bmj.b2393

References

Steyerberg, E. W., Eijkemans, M. J. C., & Habbema, J. D. F. (1999). Stepwise Selection in Small Data Sets: A Simulation Study of Bias in Logistic Regression Analysis. *Journal of clinical epidemiology*, *52*(10), 935-942. https://doi.org/10.1016/S0895-4356(99)00103-1

Stone, C. J. (1986). [Generalized Additive Models]: Comment. *Statistical science*, *1*(3), 312-314. https://doi.org/10.1214/ss/1177013607

Stone, C. J., & Koo, C.-Y. (1985). Additive splines in statistics. *Proceedings of the American Statistical Association.*, *45*, 45-49. https://www.markdiamond.com.au/download/Stone-and-Koo-1986-A4.pdf

Stovitz, S. D., & Shrier, I. (2019). Causal inference for clinicians. *BMJ evidence-based medicine*. https://doi.org/10.1136/bmjebm-2018-111069

Suzuki, E., Shinozaki, T., & Yamamoto, E. (2020). Causal diagrams: pitfalls and tips. *Journal of epidemiology*, *30*(4), 153-162. https://doi.org/10.2188/jea.JE20190192

Sylvestre, M.-P., & Abrahamowicz, M. (2008). Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in medicine*, *27*(14), 2618-2634. https://doi.org/10.1002/sim.3092

Theisen, D., Frisch, A., Malisoux, L., Urhausen, A., Croisier, J.-L., & Seil, R. (2013). Injury risk is different in team and individual youth sport. *Journal of science and medicine in sport*, *16*(3), 200-204. https://doi.org/10.1016/j.jsams.2012.07.007

Theron, G. F. (2020). *The use of Data Mining for Predicting Injuries in Professional Football Players* [Master thesis, University of Oslo]. http://urn.nb.no/URN:NBN:no-81831

Thornton, H. R., Delaney, J. A., Duthie, G. M., & Dascombe, B. J. (2017). Importance of various training-load measures in injury incidence of professional rugby league athletes. *International journal of sports physiology and performance*, *12*(6), 819-824. https://doi.org/10.1123/ijspp.2016-0326

Thoseby, B., Govus, A., Clarke, A., Middleton, K., & Dascombe, B. (2022). Temporal distribution of peak running demands relative to match minutes in elite football [journal article]. *Biology of sport*, *39*(4), 985-994. https://doi.org/10.5114/biolsport.2022.110745

Tierney, N. (2017). visdat: Visualising whole data frames. *Journal of Open Source Software*, *2*(16), 355. https://doi.org/10.21105/joss.00355

Toresdahl, B. G., Metzl, J. D., Kinderknecht, J., McElheny, K., de Mille, P., Quijano, B., & Fontana, M. A. (2022). Training patterns associated with injury in New York City Marathon runners. *British journal of sports medicine*, bjsports-2022-105670. https://doi.org/10.1136/bjsports-2022-105670

Touloumis, A. (2016). Simulating Correlated Binary and Multinomial Responses under Marginal Model Specification: The SimCorMultRes Package. *The R Journal*, *8*(2), 79-91. https://doi.org/10.32614/RJ-2016-034

Tysoe, A., Moore, I. S., Ranson, C., McCaig, S., & Williams, S. (2020). Bowling loads and injury risk in male first class county cricket: Is 'differential load' an alternative to the acute-to-chronic workload ratio? *Journal of science and medicine in sport*, *23*(6), 569-573. https://doi.org/10.1016/j.jsams.2020.01.004

UCLA. (2021). *Missing values | STATA learning modules*. UCLA: Statistical Consulting Group. Retrieved 2021-03-24 from https://stats.idre.ucla.edu/stata/modules/missing-values/

References

Udby, C. L., Impellizzeri, F. M., Lind, M., & Nielsen, R. Ø. (2020). 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*, *50*(10), 538-548. https://doi.org/10.2519/jospt.2020.9766

Ullah, S., Gabbett, T. J., & Finch, C. F. (2014). Statistical modelling for recurrent events: an application to sports injuries. *British journal of sports medicine*, *48*(17), 1287-1293. https://doi.org/10.1136/bjsports-2011-090803

Ulrich, J. (2020). *TTR: Technical Trading Rules*. In (Version 0.23) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=TTR

Van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman and Hall/CRC. https://stefvanbuuren.name/fimd/

van der Ploeg, T., Austin, P. C., & Steyerberg, E. W. (2014). Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC medical research methodology*, *14*(1), 137. https://doi.org/10.1186/1471-2288-14-137

van Ginkel, J. R., Linting, M., Rippe, R. C., & van der Voort, A. (2020). Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *Journal of personality assessment*, *102*(3), 297-308. https://doi.org/10.1080/00223891.2018.1530680

van Mechelen, W., Hlobil, H., & Kemper, H. C. G. (1992). Incidence, Severity, Aetiology and Prevention of Sports Injuries. *Sports Medicine*, *14*(2), 82-99. https://doi.org/10.2165/00007256-199214020-00002

van Smeden, M., de Groot, J. A., Moons, K. G., Collins, G. S., Altman, D. G., Eijkemans, M. J., & Reitsma, J. B. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC medical research methodology*, *16*(1), 163. https://doi.org/10.1186/s12874-016-0267-3

Vanrenterghem, J., Nedergaard, N. J., Robinson, M. A., & Drust, B. (2017). Training Load Monitoring in Team Sports: A Novel Framework Separating Physiological and Biomechanical Load-Adaptation Pathways. *Sports Medicine*, *47*(11), 2135-2142. https://doi.org/10.1007/s40279-017-0714-2

Varley, M. C., Fairweather, I. H., & Aughey1, Robert J. (2012). Validity and reliability of GPS for measuring instantaneous velocity during acceleration, deceleration, and constant motion. *Journal of sports sciences*, *30*(2), 121-127. https://doi.org/10.1080/02640414.2011.627941

Vaughan, D. (2021). *slider: Sliding Window Functions*. In (Version 0.2.1) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=slider

Verheul, J., Nedergaard, N. J., Vanrenterghem, J., & Robinson, M. A. (2020). Measuring biomechanical loads in team sports – from lab to field. *Science and Medicine in Football*, *4*(3), 246-252. https://doi.org/10.1080/24733938.2019.1709654

Vink, G. (2016). *Towards a standardized evaluation of multiple imputation routines*. Retrieved 2022-11-07 from https://www.gerkovink.com/published/%5Bworking%5D%20Towards%20a%20standardized%20evaluation%20of%20multiple%20imputation%20routines.pdf

Visnes, H., Fenstad, A. M., Stenvik, S., & Dybvik, E. (2022). ***Nasjonalt Korsbåndsregister*** *Årsrapport for 2021 med plan for forbedringstiltak*. https://www.kvalitetsregistre.no/sites/default/files/2022-09/%C3%85rsrapport%202021%20Nasjonalt%20Korsb%C3%A5ndregister.pdf

References

Von Hippel, P. T. (2009). 8. How to impute interactions, squares, and other transformed variables. *Sociological methodology*, *39*(1), 265-291. https://doi.org/10.1111/j.1467-9531.2009.01215.x

Wang, A., Healy, J., Hyett, N., Berthelot, G., & Okholm Kryger, K. (2021). A systematic review on methodological variation in acute: chronic workload research in elite male football players. *Science and Medicine in Football*, *5*(1), 18-34. https://doi.org/10.1080/24733938.2020.1765007

Wang, C., Stokes, T., Steele, R., & Shrier, I. (2021). 118 Application of the acute:chronic workload ratio in children. *British journal of sports medicine*, *55*(Suppl 1), A47-A48. https://doi.org/10.1136/bjsports-2021-IOC.110

Wang, C., Stokes, T., Vargas, J. T., Steele, R., Wedderkopp, N., & Shrier, I. (2021). Injury risk increases minimally over a large range of the acute-chronic workload ratio in children. *American journal of epidemiology*. https://doi.org/10.1093/aje/kwab280

Wang, C., Stokes, T., Vargas, J. T., Steele, R., Wedderkopp, N., & Shrier, I. (2022). Predicting Injury Risk Over Changes in Physical Activity in Children Using the Acute:Chronic Workload Ratio. *American journal of epidemiology*, *191*(4), 665-673. https://doi.org/10.1093/aje/kwab280

Wang, C., Vargas, J. T., Stokes, T., Steele, R., & Shrier, I. (2020). Analyzing Activity and Injury: Lessons Learned from the Acute: Chronic Workload Ratio. *Sports Medicine*, 1-12. https://doi.org/10.1007/s40279-020-01280-1

Wang, C. S., Tyrel, Steele, R., Wedderkopp, N., & shrier, I. (2020). Implementing multiple imputation for missing data in longitudinal studies when models are not feasible: A tutorial on the random hot deck approach. *arXiv preprint methdology*. https://doi.org/arXiv:2004.06630v4

Ward, P., Coutts, A. J., Pruna, R., & McCall, A. (2018). Putting the "I" Back in Team. *International journal of sports physiology and performance*, *13*(8), 1107-1111. https://doi.org/10.1123/ijspp.2018-0154

Warren, A., Williams, S., McCaig, S., & Trewartha, G. (2018). High acute: chronic workloads are associated with injury in England & Wales Cricket Board Development Programme fast bowlers. *Journal of science and medicine in sport*, *21*(1), 40-45. https://doi.org/10.1016/j.jsams.2017.07.009

Weaving, D., Dalton-Barron, N., McLaren, S., Scantlebury, S., Cummins, C., Roe, G., Jones, B., Beggs, C., & Abt, G. (2020). The relative contribution of training intensity and duration to daily measures of training load in professional rugby league and union. *Journal of sports sciences*, *38*(14), 1674-1681. https://doi.org/10.1080/02640414.2020.1754725

Weiss, K. J., Allen, S. V., McGuigan, M. R., & Whatman, C. S. (2017). The relationship between training load and injury in men's professional basketball. *International journal of sports physiology and performance*, *12*(9), 1238-1242. https://doi.org/10.1123/ijspp.2016-0726

West, S., Williams, S., Cazzola, D., Cross, M., Kemp, S., & Stokes, K. (2021). 119 The acute:chronic workload ratio: why one size does not fit all. *British journal of sports medicine*, *55*(Suppl 1), A48-A48. https://doi.org/10.1136/bjsports-2021-IOC.111

West, S. W., Clubb, J., Torres-Ronda, L., Howells, D., Leng, E., Vescovi, J. D., Carmody, S., Posthumus, M., Dalen-Lorentsen, T., & Windt, J. (2021). More than a Metric: How

Training Load is Used in Elite Sport for Athlete Management. *International journal of sports medicine*, *42*(04), 300-306. https://doi.org/10.1055/a-1268-8791

West, S. W., Williams, S., Cazzola, D., Kemp, S., Cross, M. J., & Stokes, K. A. (2020). Training load and injury risk in elite rugby union: the largest investigation to date. *International journal of sports medicine*. https://doi.org/10.1055/a-1300-2703

Weston, S., & Microsoft. (2022a). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. In (Version 1.0.17) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=doParallel

Weston, S., & Microsoft. (2022b). *foreach: Provides Foreach Looping Construct*. In (Version 1.5.2) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=foreach

Westreich, D., & Greenland, S. (2013). The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American journal of epidemiology*, *177*(4), 292-298. https://doi.org/10.1093/aje/kws412

Wheeler, D. J., & Chambers, D. S. (2010). *Understanding Statistical Process Control* (3rd ed.). SPC Press. https://www.spcpress.com/book_understanding_statistical_process_control.php

White, I. R., & Carlin, J. B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in medicine*, *29*(28), 2920-2931. https://doi.org/10.1002/sim.3944

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*, *30*(4), 377-399. https://doi.org/10.1002/sim.4067

Wickham, H. (2019). *Tidyverse*. Retrieved 2022-11-07 from https://www.tidyverse.org/

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., & Hester, J. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilkinson, M., & Akenhead, R. (2013). Violation of Statistical Assumptions in a Recent Publication? *International journal of sports medicine*, *34*(03), 281. https://doi.org/10.1055/s-0032-1331775

Williams, S., Trewartha, G., Cross, M. J., Kemp, S. P., & Stokes, K. A. (2017). Monitoring what matters: a systematic process for selecting training-load measures. *International journal of sports physiology and performance*, *12*(s2), S2-101-S102-106. https://doi.org/10.1123/ijspp.2016-0337

Williams, S., Trewartha, G., Kemp, S. P., Brooks, J. H., Fuller, C. W., Taylor, A. E., Cross, M. J., & Stokes, K. A. (2016). Time loss injuries compromise team success in Elite Rugby Union: a 7-year prospective study. *British journal of sports medicine*, *50*(11), 651-656. https://doi.org/10.1136/bjsports-2015-094798

Williams, S., West, S., Cross, M. J., & Stokes, K. A. (2017). Better way to determine the acute: chronic workload ratio? *British journal of sports medicine*, *51*(3), 209-210. https://doi.org/10.1136/bjsports-2016-096589

Windt, J., Ardern, C. L., Gabbett, T. J., Khan, K. M., Cook, C. E., Sporer, B. C., & Zumbo, B. D. (2018). Getting the most out of intensive longitudinal data: a methodological review of workload–injury studies. *BMJ open*, *8*(10), e022626. https://doi.org/10.1136/bmjopen-2018-022626

Windt, J., & Gabbett, T. J. (2017). How do training and competition workloads relate to injury? The workload—injury aetiology model. *British journal of sports medicine*, *51*(5), 428-435. https://doi.org/10.1136/bjsports-2016-096040

Windt, J., Gabbett, T. J., Ferris, D., & Khan, K. M. (2017). Training load--injury paradox: is greater preseason participation associated with lower in-season injury risk in elite rugby league players? *British journal of sports medicine*, *51*(8), 645-650. https://doi.org/10.1136/bjsports-2016-095973

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama*, *310*(20), 2191-2194.

Xiao, M., Nguyen, J. N., Hwang, C. E., & Abrams, G. D. (2021). Increased Lower Extremity Injury Risk Associated With Player Load and Distance in Collegiate Women's Soccer. *Orthopaedic journal of sports medicine*, *9*(10). https://doi.org/10.1177/23259671211048248

Xiao, X., White, E. P., Hooten, M. B., & Durham, S. L. (2011). On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. *Ecology*, *92*(10), 1887-1894. https://doi.org/10.1890/11-0538.1

Zeileis, A., & Grothendieck, G. (2005). zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, *14*(6), 1–27. https://doi.org/10.18637/jss.v014.i06

Zouhal, H., Boullosa, D., Ramirez-Campillo, R., Ali, A., & Granacher, U. (2021). Acute: Chronic Workload Ratio: Is There Scientific Evidence? *Frontiers in physiology*, *12*. https://doi.org/10.3389/fphys.2021.66968

Zumeta-Olaskoaga, L., Weigert, M., Larruskain, J., Bikandi, E., Setuain, I., Lekue, J., Küchenhoff, H., & Lee, D.-J. (2021). Prediction of sports injuries in football: a recurrent time-to-event approach using regularized Cox models. *AStA Advances in Statistical Analysis*. https://doi.org/10.1007/s10182-021-00428-2

7 Appendices

Appendix I: Papers

Paper I

Science and Medicine in Football

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rsmf20

Handling and reporting missing data in training load and injury risk research

L. K. Bache-Mathiesen, Thor Einar Andersen, Benjamin Clarsen & Morten Wang Fagerland

To cite this article: L. K. Bache-Mathiesen, Thor Einar Andersen, Benjamin Clarsen & Morten Wang Fagerland (2021): Handling and reporting missing data in training load and injury risk research, Science and Medicine in Football, DOI: 10.1080/24733938.2021.1998587

To link to this article: https://doi.org/10.1080/24733938.2021.1998587 https://doi.org/10.1080/24733938.2021.1998587

Routledge
Taylor & Francis Group

Check for updates

Handling and reporting missing data in training load and injury risk research

L. K. Bache-Mathiesen [a], Thor Einar Andersen [a], Benjamin Clarsen [a,b] and Morten Wang Fagerland [a,c]

[a]Oslo Sports Trauma Research Centre, Department of Sports Medicine, Norwegian School of Sports Sciences, Oslo, Norway; [b]Centre for Disease Burden, Norwegian Institute of Public Health, Bergen, Norway; [c]Oslo Centre for Biostatistics and Epidemiology, Research Support Services, Oslo University Hospital, Oslo, Norway

ABSTRACT

Purpose: To map the current practice of handling missing data in the field of training load and injury risk and to determine how missing data in training load should be handled.

Methods: A systematic review of the training load and injury risk literature was performed to determine how missing data are reported and handled. We ran simulations to compare the accuracy of modelling a predetermined relationship between training load and injury risk following handling missing data with different methods. The simulations were based on a Norwegian Premier League men's football dataset (n = 39). Internal training load was measured with the session Rating of Perceived Exertion (sRPE). External training load was the total distance covered measured by a global positioning systems (GPS) device.

Results: Only 37 (34%) of 108 studies reported whether training load had any missing observations. Multiple Imputation using Predicted Mean Matching was the best method of handling missing data across multiple scenarios.

Conclusion: Studies of training load and injury risk should report the extent of missing data, and how they are handled. Multiple Imputation with Predicted Mean Matching should be used when imputing sRPE and GPS variables.

ARTICLE HISTORY
Accepted 19 October 2021

KEYWORDS
Training load; injury; missing data; simulation; systematic review

Introduction

Sport injuries are detrimental to athlete health and performance (Hägglund et al. 2013), and are a considerable cost for clubs and sport organizations (Fernández-Cuevas et al. 2010). With the ultimate goal of injury prevention, researchers in sports and medicine science strive to identify risk factors for injury (Bahr and Krosshaug 2005). One potential, modifiable risk factor is training load (Windt and Gabbett 2017). To assess the relationship between training load and injury risk, researchers have often collected longitudinal sports data and performed regression modelling (Windt et al. 2018). The ability of such analyses to determine whether training load affects the risk of injury, and the level of certainty surrounding the estimates, depends on the study design (Lang 2005; Shmueli 2010), statistical choices (Lang 2004; Mansournia et al. 2021), and sample size (statistical power, Bahr and Holme 2003). So far, sample sizes in the field of training load and injury research have been criticized (Griffin et al. 2020; Andrade et al. 2020).

A handful of athletes may have hundreds of training load values each (De Leeuw et al. 2021). When analysing the relationship between training load and injury risk, the main factor affecting statistical power is the number of injuries (Bahr and Holme 2003). In the field of sports injuries, associations are often small to moderate (Bahr 2016), which requires larger sample sizes (number of events) than strong associations. A sample size calculation in Bahr and Holme (2003) suggested at least 200 injuries. While determining required number of events is too complex to boil down to a simple rule of thumb (Van Smeden et al. 2016; Riley et al. 2019; Nielsen et al. 2019), the number needed is likely to be higher than currently seen in training load and injury risk studies (Griffin et al. 2020). Therefore, it is critical to retain as many injuries as possible.

In longitudinal data collection, missing data is almost inevitable (Karahalios et al. 2012). In Enright et al. (2019), injuries (53%) were excluded from analyses due to inconsistent and/or missing data in training load. Similarly, in a 3-season football cohort, 124 (81%) out of 154 eligible injuries were excluded due to insufficient training load data (Lolli et al. 2020). Therefore, missing observations in training load, unless dealt with appropriately, introduce missing injury data. Missing training load data may, depending on the mechanism for missing data, also introduce selection bias. Practices such as removing athletes transferred to other clubs (Moreno-Pérez et al. 2020), including only those who completed > 80% of surveys (Theisen et al. 2013; Albrecht et al. 2020), or those who completed a full season (Fanchini et al. 2018), remove consenting participants with partial data and reduces the external validity of results. There is, however, no consensus on how to handle missing data in training load research (Mccall et al. 2018). Ideally, missing data should be handled in a way that retains the properties of the observed data and does not affect study conclusions – as though there were no missing data to begin with.

Therefore, the purpose of this study was to determine the best methods of handling missing data in training load and injury research. First, we performed a systematic review of the training load and injury risk literature to map current practices

CONTACT L. K. Bache-Mathiesen l.k.bache-mathiesen@nih.no Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo 0806, Norway
Supplemental data for this article can be accessed here.

for handling missing data and to guide our choices of methods. We then compared the ability of different methods of imputing and deleting missing observations to accurately model a known relationship between training load and injury risk. Based on our results, we propose a best-practice guide to reporting and handling missing data in training load research.

Materials and methods

Systematic review of missing data in training load and injury studies

For an estimate of the current practice of reporting and handling missing data, we performed a systematic review of the field of training load and injury risk. Inclusion criteria were as follows:

- Studied a sports population
- Analysed training load or relative training load as the exposure of interest
- Analysed injury or health problem risk as the outcome
- For a picture of the latest decade, year of publication ≥ 2010

Studies were extracted from the most recent, relevant reviews (Eckard et al. 2018; Windt et al. 2018; Griffin et al. 2020; Maupin et al. 2020; Andrade et al. 2020; Udby et al. 2020; Dalen-Lorentsen et al. 2021). To achieve a more accurate picture of current reporting practices, the PubMed database was also searched for training load and injury risk studies published between 2019 and 2021. The search terms were 'training load AND injury', 'workload AND injury', 'ACWR', and 'acute: chronic'. The search yielded 125 studies (Bache-Mathiesen 2021b). Of these, 17 were published before 2010 and excluded from the analyses.

We calculated the proportion of studies reporting whether they had missing observations in the training load measure, by year and overall. For those with missing data, we determined the mean amount of missing observations in the training load variable and the methods used to handle missing data. We used our findings to determine which methods should be compared in the simulation.

Comparison of methods for handling missing data

To compare the performance of different statistical approaches, it is common to run a simulation (Morris et al. 2019). In our study, we constructed a relationship between training load and injury using real, observed training load measures from a football dataset. In the dataset with this known relationship between training load and injury, we deleted different amounts of training load observations. The methods we wished to compare were used to impute or delete the missing training load data. When the same analysis was used to determine the relationship between training load and injury – the only difference from one analysis to another was the choice of method for handling missing data – the amount of deviation from the known relationship could be measured, and the methods compared (Vink 2016).

Observed sports data

Participants

A total of 42 male professional football players from a Norwegian Premier League team (mean age 26 years, standard deviation (SD): 4) were followed for 323 days during the 2019 season (Theron 2020).

The study was approved by the Ethical Review Board of the Norwegian School of Sport Sciences, and the Norwegian Centre for Research Data (722773). All participants provided informed written consent.

Training load definition

Session rating of perceived exertion. Daily, within 30 minutes after completion of each training session or match (Rønneberg 2020), the players reported the duration of each sporting activity and their internal load expressed as Rating of Perceived Exertion (RPE, Borg et al. 1987) on the Foster et al. (2001) scale, using a mobile application (Athlete Monitoring, Moncton, Canada). For each activity, the RPE was multiplied by the duration of the activity in minutes to derive the session RPE (sRPE, Foster et al. 2001).

Global positioning systems. Global Positioning Systems (GPS) were used to collect external training load measures (Bourdon et al. 2017), with 10 Hz sampling rate (Catapult OptimEye X4, Catapult Sports, Australia). Each player always used the same device, which was applied 30 minutes prior to start of the training sessions and matches (Rønneberg 2020). Exported variables included (1) total distance covered, (2) distance covered above 20 km/h (high-speed running distance), (3) distance covered above 25 km/h (sprint distance), and (4) the squared instantaneous rate of change in acceleration for three vectors of direction (x, y, and z axes) divided by 100 (player load, Boyd et al. 2011).

As there is no consensus in the literature on definitions of high-speed or sprint-speed measured by GPS devices (Dwyer and Gabbett 2012), and only total distance had the same definition across training load and injury risk studies (Andrade et al. 2020; Maupin et al. 2020), it was the main focus of our study.

Additional variables. Table 1 shows all variables included in the study. Additional variables included player ID (anonymised), date of activity, whether the load was during a match (yes/no), and the micro-cycle-day (M, M-1, M-2, M-3, M-4, M + 2, M + 1). A micro-cycle consisted of all activity before a new match (M, Bache-Mathiesen et al. 2021). That is, recovery days after the previous match as well as the training days before the next match. Days denoted with negative numbers are training days before the next match (M-1; being the day before the match, M-2; two days before a match, and so on). Days with positive numbers are recovery and training days after a match (M + 1; being the day after a match, M + 2; two days after a match). On recovery days, players only reported activity parameters if they participated in an activity, and so, if total distance, RPE or session duration was missing on M + 1 or M + 2 days, they were assumed to be 0.

Table 1. Overview of variables included in the study.

Dataset	Variable	Type	Units
sRPE	Player ID	Nominal	Integer
	Date of activity	Date	Year-Month-Day (YYYY-MM-DD)
	Match	Logical	Yes/No
	Micro-cycle-day[a]	Nominal	M, M-1, M-2, M-3, M-4, M + 2, M + 1
	RPE	Continuous	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
	Minutes in activity	Continuous	Minutes
	Playing Position[b]	Nominal	Central Defender, Fullback, Central Midfielder, Winger, Striker, Goal Keeper
GPS	Player ID	Nominal	Integer
	Date of activity	Date	Year-Month-Day (YYYY-MM-DD)
	Match	Logical	Yes/No
	Micro-cycle-day[a]	Nominal	M, M-1, M-2, M-3, M-4, M + 2, M + 1
	Total distance	Continuous	m/day
	High-speed running distance	Continuous	m/day
	Sprint distance	Continuous	m/day
	Player load	Continuous	m/day
	Playing Position[b]	Nominal	Central Defender, Fullback, Central Midfielder, Winger, Striker, Goal Keeper
	sRPE[b]	Continuous	Minutes*RPE (Arbitrary Units)

Abbreviations: RPE, Rating of Perceived Exertion; sRPE, session Rating of Perceived Exertion; m/day, meters per day

[a]Days denoted with negative numbers are training days before the next match (M-1; being the day before the match, M-2; two days before a match, and so on). Days with positive numbers are recovery and training days after a match (M + 1; being the day after a match, M + 2; two days after a match).

[b]These variables were only available in select analyses

Simulations

Comparison of imputation methods

We performed stochastic simulations to compare different methods of imputing and deleting missing data in training load before modelling the relationship between training load and injury risk. A detailed description of the simulation process and equations, as well as justifications for our methodological choices, is available as an appendix (supplemental file I). See Box 1 for a summary of the simulation steps.

Box 1. A summary of the simulation steps.

(1) Add missing drawn under MCAR and MAR from the original dataset.
(2) Impute or delete missing data using five different methods.
(3) Fit logistic regression models with injury as the outcome and training load as the explanatory variable on the imputed or missing-omitted data.

Steps 1–3 was repeated 1 900 times for acceptable accuracy according to a sample size calculation (Morris et al. 2019).

First, all missing observations was removed from all variables in the datasets. The final datasets had 4 782 sRPE values and 2 292 total distance values of Gaussian distributions (supplemental file II Figure S1). Simulated injuries were added to the datasets with a predefined, linear relationship between training load and the probability of injury. This resulted in 1 333 and 859 injuries in the two datasets, respectively. Logistic regression was run with injury as the response variable and training load as the explanatory variable to determine performance when no data were missing. We caution performing such analyses in a real study; it is unreasonable to assume the relationship between training load and injury risk is as simple as in this simulation, and more advanced methods for dealing with repeated measures and recurrent events are needed (Nielsen et al. 2020).

Imputation strategy for a derived variable

It is unknown how a derived variable such as sRPE should be imputed (Van Buuren 2018; Benson et al. 2021). Simulations were performed to compare four strategies of imputing sRPE

- Impute, then transform (Von Hippel 2009). In this method, the product (sRPE) is not available to inform the imputation model. However, it may reduce collinearity issues.
- Transform, then impute (Von Hippel 2009; White et al. 2011). Here, sRPE is present in the imputation model.
- Passive imputation (Van Buuren 2018). The relationship between intensity, duration and sRPE is described in the imputation model, which may be an improvement over using them merely as explanatory variables.
- Impute product without factors. Under this scenario, no issues stemming from the strong correlation between intensity, duration and sRPE are present, but intensity and duration are not available to inform the imputation model.

The most accurate method determined in these simulations were used in Step 2 below.

Step 1 add missing

We hypothesized that missing observations in total distance and other GPS variables may often be Missing Completely at Random (MCAR). Under MCAR, all observations have an equal probability of missing – the probability is not dependent on other factors. For instance, we can imagine that technical errors in a GPS device can happen at random, and are not dependent on the characteristics of the athlete or the performed activity.

For RPE – an athlete-reported parameter – we theorized that it is more likely that the probability of missing data depends on characteristics of the player and also of the day of activity, which was the case in Benson et al. (2021). This assumption is known as Missing at Random (MAR, Janssen et al. 2010). As a hypothetical example, players may be busier on match days and forget to report RPE. In such a scenario, whether a day is a match can predict the probability of missing RPE data. Missing GPS data may also be MAR if, for instance, the devices were worn in different locations and environmental obstruction was present in some locations and not in others (Malone et al. 2017).

Figure 1. Illustration of the modeling process in the framework of multiple imputation. In the first step, the variables available in the dataset are used to predict *m* number of potential imputations for the missing observations. A dataset is created for each of the *m* sets of predictions. The number of imputed datasets, five, is recommended in most cases (Van Buuren 2018, section 2.8). The main model of interest is then run on each of the 5 datasets. The estimates from each model are averaged using Rubin's rules, which calculate standard errors that account for between-imputation variation and the level of uncertainty that stems from the missing data (Sterne et al. 2009; Van Buuren 2018).

From the sRPE dataset and the total distance dataset, eleven datasets were created with different amounts of missing sampled under the assumption of MCAR, and likewise, three under the assumption of MAR, where correlations were increasingly stronger between variables and the probability of missing. This ensured covering a range of missing data amounts and mechanisms (Vink 2016; Schouten and Vink 2018).

Step 2 impute or delete missing

We imputed or deleted the missing observations in the fourteen sRPE and the fourteen total distance datasets created in Step 1 with five different methods, respectively:

- Complete Case Analysis (Listwise deletion, White and Carlin 2010).
- Mean Imputation using the player mean (Benson et al. 2021; Wang et al. 2021).
- Mean Imputation using the weekly mean (Benson et al. 2021).
- Multiple Imputation using Predicted Mean Matching (Figure 1, Van Buuren 2018).
- Regression Imputation (Musil et al. 2002).

Step 3 fitting models on imputed data

We ran logistic regression models with training load as the independent variable, and the simulated injuries as the outcome variable, on each of the fourteen sRPE and fourteen total distance datasets imputed in Step 2.

Performance measures

We used the following performance measures (Van Buuren 2018) to compare the performance of the different imputation methods:

- Percent bias (PB). The upper limit for acceptable performance was ±5% (Demirtas et al. 2008).
- Root-Mean-Squared Error (RMSE). If all methods have acceptable bias, they may be distinguished by RMSE.
- Coverage: the proportion of 95% confidence intervals that contained the true value.
- Average width (AW) of the 95% confidence interval. If all methods have a coverage > 95%, they may be distinguished by average width.

The PB per method was visualized for each scenario of missing. In addition, a visualization of imputed versus observed data was created.

Imputation with extra variables available

For more realistic missing data scenarios (Schouten et al. 2018), the simulations were repeated to test whether the results changed with the inclusion or exclusion of the player's playing position in the imputation model. We also tested whether the results changed if all GPS-variables were missing whenever "total distance", the focus of our study, was missing.

Results

Systematic review of training load and injury studies

The characteristics of 108 studies that assessed the relationship between training load and injury risk are reported in Table 2. A total of 37 (34%) studies reported whether the training load variable had any missing observations, between 30%–50% the last five years (Figure 2). Of these, 25 (23%) studies described how missing data were handled (Table 3). The most popular methods were Mean Imputation (n = 11) and Complete Case Analysis (n = 8). For 18 studies that reported the amount of missing data in the training load variable, the Mean Percentage Missing was 7.3% (SD = 6%).

In 9 (24%) of the 37 studies reporting missing data, athletes were removed due to incomplete or missing data, and in 7 (19%) studies, injuries were removed. Overall, the mean percentage of removed athletes was 13% (SD = 10), and the

Table 2. The characteristics of the N = 108 studies assessing the relationship between training load and injury risk.

Study Characteristic	N studies	% of studies
Sex		
Males only	85	79%
Males and Females	20	19%
Females only	3	3%
Training load measure[a]		
sRPE	41	38%
GPS	27	25%
sRPE and GPS	16	15%
Time (hours/minutes in activity)	8	7%
Other measures	16	15%
Study period		
1 season/year/school year	48	52%
2 seasons/years/school years	18	19%
3 seasons/years/school years	6	7%
4 seasons/years/school years	7	8%
≥ 5 seasons/years/school years	4	4%
Other study length metrics[b]	10	11%
Sport		
Football (soccer)	29	27%
Australian Football	20	19%
Rugby	15	14%
Cricket	9	8%
Endurance Sports	7	7%
Multiple Sports	6	6%
Gaelic Football	4	4%
American Football	3	3%
Basketball	3	3%
Tennis	3	3%
Volleyball	3	3%
Handball	2	2%
Other sports[c]	4	4%

[a]sRPE = session Rating of Perceived Exertion; GPS = Global Positioning System; Other measures = sport specific measures such as 'balls bowled' in cricket, or 'number of jumps' in volleyball; Heart-Rate monitoring; Match Exposure and more.

[b]'Other study length metrics' encompasses reports of 1 preseason, 1 competition and 1 training camp.

[c]'Other sports' encompasses alpine ski racing, baseball, CrossFit and hurling.

mean percentage of removed injuries was 34% (SD = 30). The mean number of injuries analysed in the studies was 210 (SD = 703) and the median 85; 81% of studies had ≤ 200 injuries (Figure 3).

Table 3. The methods used to handle missing observations in the training load variable as reported by 36[a] studies in the field of training load and injury risk research.

Missing Data Method	N studies	% of studies
Unclear[b]	12	33%
Mean Imputation	11	31%
Complete Case Analysis	8	22%
Median Imputation	2	6%
Multiple Imputation	2	6%
Regression Imputation	1	3%

[a]Of 108 eligible studies, 37 (34%) reported whether they had missing data in the training load variable. One of the 37 studies had no missing data and was removed from this analysis.

[b]Methods considered 'Unclear' were cases where authors reported having missing data, but the method used could not be determined.

Missing data in the Norwegian premier league dataset

Across 4 871 days of activity, 650 (13%) RPE observations were missing. Of 42 players, 6 (14%) provided no RPE responses. Also, 3 (7%) had no GPS data collection. The remaining players applied GPS devices for 2 984 days, of which 122 (4%) were missing due to technical errors. This number is an underestimation, as a database-programmer removed an unknown number of empty GPS files, which then could not be included in the calculation. Missing observations were concentrated towards the end of the study period. The probability of missing is likely dependent on time, and therefore deemed to be Missing at Random (supplemental file II Figure S2).

Simulations

Imputing with the Predicted Mean Matching method had a lower bias using Multiple Imputation (mean %-bias = 3.5) than Single Imputation (mean %-bias = 4.1), in addition to a higher standard error (mean SE = 0.0000218 vs. mean SE = 0.0000206, supplemental file II Table S1).

Complete Case Analysis and Multiple Imputation using Predicted Mean Matching were the only methods that retained the distribution of the observed sRPE and the total distance data under the assumption of both MCAR

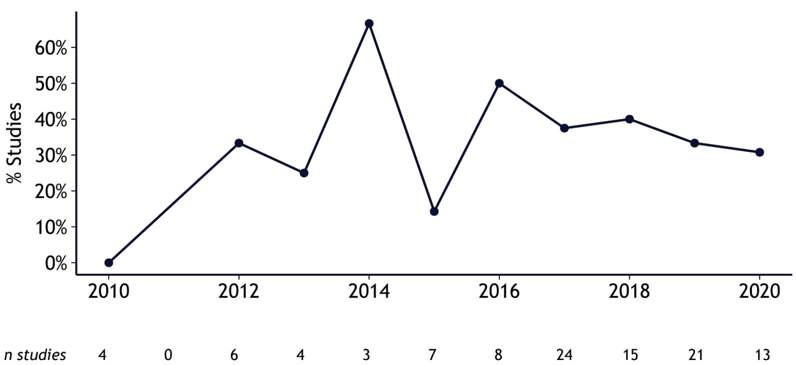

Figure 2. The percentage of studies, by year of publication, that reported whether they had missing data in training load. Since 2021 had yet to come to pass, it was not comparable to previous years, and so, the analysis was based on the 105 training load and injury risk studies published in the period 2010–2020, only. N studies are the number of studies published in each year (the denominator).

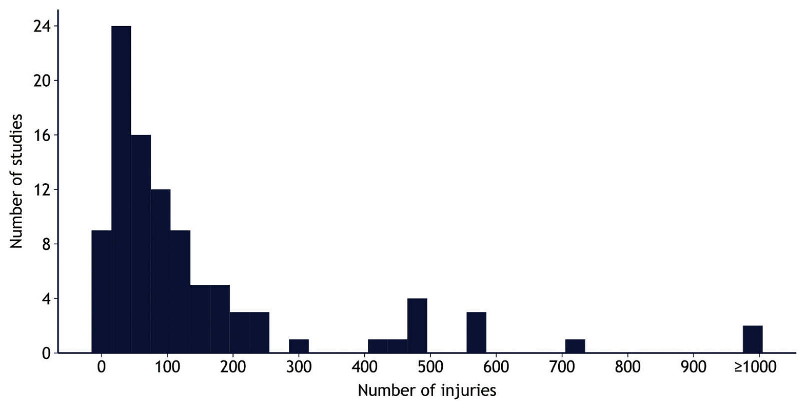

Figure 3. Distribution of the number of injuries reported in 99 studies assessing the relationship between training load and injury risk. The distribution is concentrated below 100 injuries, with a median of 85. Of 108 studies, 99 reported the number of injuries, and 80 of these had ≤ 200 injuries (81%), based on the final number of injuries used in the studies' analyses of injury risk.

Table 4. The mean performance of imputing or deleting missing observations across 11 scenarios of Missing Completely at Random (MCAR) and 3 scenarios of Missing at Random (MAR) in session Rating of Perceived Exertion (sRPE), and total distance covered for the situation where missing is introduced to all GPS variables and missing is introduced to total distance, only (N simulations = 1 900). Compared to performance without missing data (None).

Missing	Missing variables		RB[a]	PB[a]	RMSE[a]	Coverage[b]	AW
None			0.00004189	1.4%	0.0000419	100%	0.0006236
MCAR	sRPE	Complete Case Analysis	0.00030114	10%	0.0003188	95%	0.0019103
		Mean per player	0.00034171	11.4%	0.0003572	80%	0.0009377
		Mean per week	0.00031297	10.4%	0.0003376	65%	0.0009215
		MI – PMM	0.00017309	5.8%	0.0001908	95%	0.0013997
		Regression Imputation	0.00101082	33.7%	0.0010398	30%	0.0008282
MAR		Complete Case Analysis	0.00022987	7.7%	0.0002798	100%	0.0016990
		Mean per player	0.00023632	7.9%	0.0002544	100%	0.0009914
		Mean per week	0.00026460	8.8%	0.0002751	100%	0.0009678
		MI – PMM	0.00011496	3.8%	0.0001441	100%	0.0011119
		Regression Imputation	0.00116333	38.8%	0.0011751	33%	0.0008325
None			0.00001190	4%	0.0000119	100%	0.0000805
MCAR	All GPS variables	Complete Case Analysis	0.00002455	8.2%	0.0000246	100%	0.0001255
		Mean per player	0.00002574	8.6%	0.0000258	90%	0.0001027
		Mean per week	0.00003939	13.1%	0.0000394	90%	0.0000987
		MI – PMM	0.00003213	10.7%	0.0000321	87%	0.0000816
		Regression Imputation	0.00007450	24.8%	0.0000745	40%	0.0000913
	Total distance only	Complete Case Analysis	0.00002672	8.9%	0.0000267	93%	0.0001220
		Mean per player	0.00002891	9.6%	0.0000289	90%	0.0001024
		Mean per week	0.00005448	18.2%	0.0000545	70%	0.0000984
		MI – PMM	0.00000983	3.3%	0.0000098	100%	0.0000808
		Regression Imputation	0.00001820	6.1%	0.0000182	100%	0.0000776
MAR	All GPS variables	Complete Case Analysis	0.00006099	20.3%	0.0000610	78%	0.0001559
		Mean per player	0.00003401	11.3%	0.0000340	78%	0.0001229
		Mean per week	0.00003511	11.7%	0.0000351	89%	0.0001200
		MI – PMM	0.00002850	9.5%	0.0000285	78%	0.0001977
		Regression Imputation	0.00018145	60.5%	0.0001814	67%	0.0001241
	Total distance only	Complete Case Analysis	0.00004217	14.1%	0.0000422	100%	0.0001433
		Mean per player	0.00002977	9.9%	0.0000298	100%	0.0001175
		Mean per week	0.00003416	11.4%	0.0000342	89%	0.0001143
		MI – PMM	0.00002024	6.7%	0.0000202	100%	0.0000975
		Regression Imputation	0.00003498	11.7%	0.0000350	67%	0.0000841

Abbreviations: AW, Average Width of 95% confidence intervals; GPS, Global Positioning System; PB, Absolute Percent Bias; sRPE, Session Rating of Perceived Exertion; RB, Absolute Raw Bias; RMSE, Root-Mean-Squared-Error
[a]Monte Carlo standard error < 0.0001
[b]Monte Carlo standard error = 0.5

and MAR (supplemental file II Figures S3–S6). Complete Case Analysis had the highest average width of 95% confidence intervals for all scenarios of missing data (Table 4).

Logistic regression performed on the dataset without any missing data had a bias of 1.4% for sRPE and 4% for total distance (Table 4).

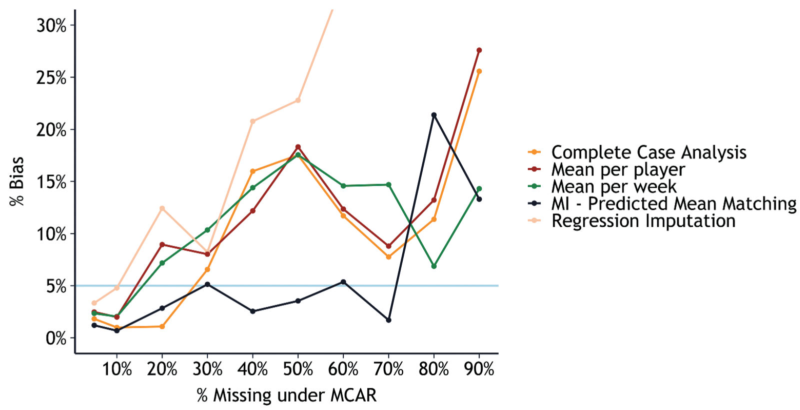

Figure 4. Performance of five different methods of imputing or deleting missing observations in the session rating of perceived exertion. performance is measured by the mean absolute percent bias (% Bias). Varying levels of missing data under the assumption of Missing Completely at Random (MCAR) are displayed along the X-axis. The best methods are closest to 0, and the light blue line indicates the maximum range of acceptable bias (0% to 5%). No method stayed consistently within acceptable bias after 50% missing. Regression imputation reached 49% bias at 80% missing and 136% bias at 90% missing, and therefore was off the chart. Logistic regression performed on the data without missing had a bias of 1.4%. Based on 1 900 simulations with monte carlo standard error < 0.00001.

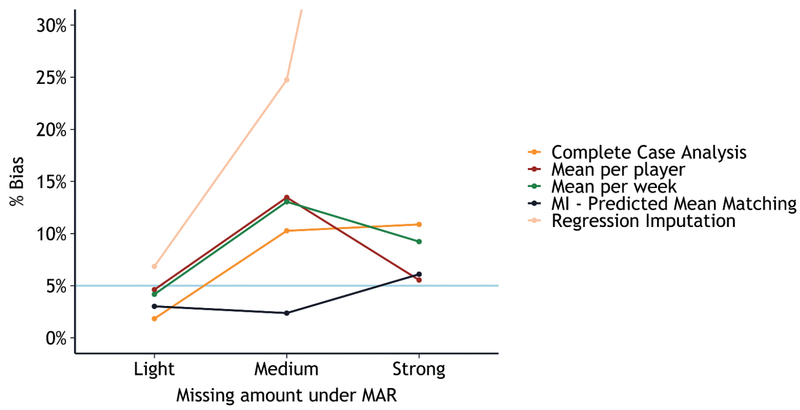

Figure 5. Performance of five different methods of imputing or deleting missing observations in the session rating of perceived exertion. performance is measured by the mean absolute percent bias (% Bias). Levels of missing data were added under the assumption of Missing at Random (MAR): Light MAR (≈25% missing); Medium MAR (≈50% missing); Strong MAR (≈80% missing). Under MAR, the probability of missing is dependent on other variables. The best methods are closest to 0, and the light blue line indicates the maximum range of acceptable bias (0% to 5%). Multiple Imputation (MI) using Predicted Mean Matching was the only method with consistently acceptable bias through Medium MAR. Regression imputation reached off the chart with ≈85% under strong MAR. Logistic regression performed on the data without missing had a bias of 1.4%. Based on 1 900 simulations with monte carlo standard error < 0.00001.

Handling missing data in session rating of perceived Exertion

In the comparison of different orders of imputing sRPE as a derived variable, the method 'Impute, then transform' had the lowest bias (1.4%), and highest certainty compared to other methods (supplemental file II Table S2). Thus, in the main simulation, missing data was imputed for RPE and duration, and sRPE was calculated afterwards.

Complete Case Analysis and Multiple Imputation using Predicted Mean Matching were the only methods within acceptable bias (< |5%|) consistently up to 20% missing sRPE observations under MCAR (Figure 4). Predicted Mean Matching was subsequently within acceptable bias up to 50% missing and had the lowest bias on average (6% vs. ≥ 10% [all other methods], Table 4) and an adequate coverage of 95% (Table 4).

Under MAR, Multiple Imputation using Predicted Mean Matching was within acceptable bias up to and including ≈50% missing (3% bias at ≈25% missing, 2.4% at ≈50% missing, Figure 5) with good coverage (100%, Table 4). Adding the player's playing position to the imputation model improved performance under MCAR but not under MAR (supplemental file II Figure S7).

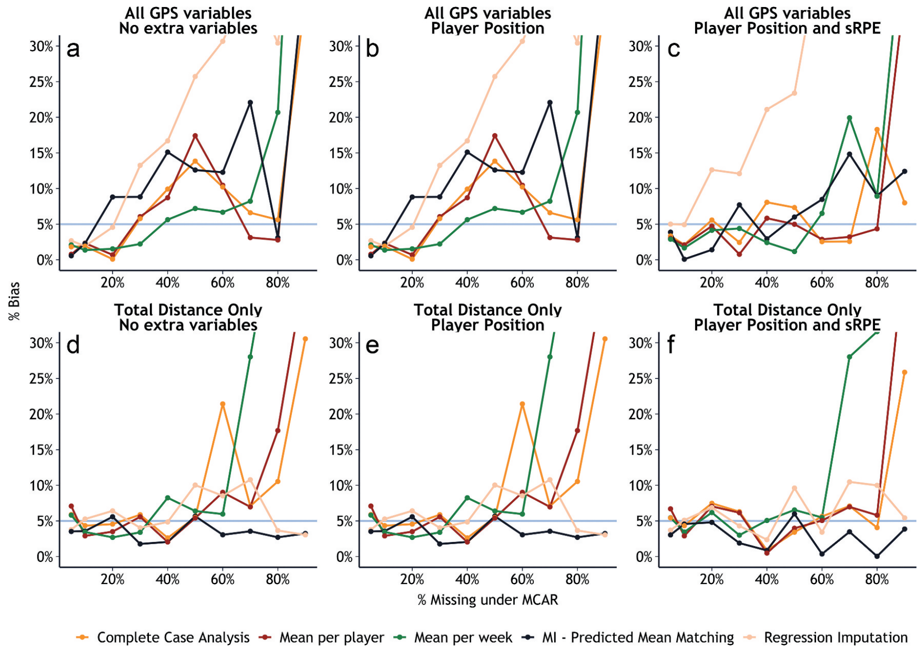

Figure 6. Performance of five different methods of imputing or deleting missing observations in training load measured by Global Positioning Systems (GPS) variables. Performance is measured by the Mean Percent Bias (% Bias). Varying levels of missing data under the assumption of Missing Completely at Random (MCAR) are displayed along the X-axis. For (a–c), missing is introduced to all four GPS variables; (d–f) missing is only introduced to the total distance variable. In addition, for (a, d) there were no extra variables in the imputation model, (b, e) the player position was in the imputation model, and (c, f) the session Rating of Perceived Exertion (sRPE) and the player position was in the imputation model. The best methods are closest to 0, and the light blue line indicates the maximum range of acceptable bias (0% to 5%). MI = Multiple Imputation. Methods off the chart had > 30% bias. Logistic regression performed on the data without missing had a bias of 4%. Based on 1 900 simulations with monte carlo standard error < 0.00001.

Handling missing data in total distance

When missing data was in the total distance variable, only – not in other GPS variables – Complete Case Analysis and Multiple Imputation using Predicted Mean Matching had acceptable bias for less than 50% missing under MCAR (Figure 6(d–f)). Predicted Mean Matching was the only method of acceptable bias for amounts of missing data from 50%, up to 90% (mean bias 3.3%, 100% coverage, Table 4). However, when all GPS variables were missing – high-speed running distance, sprint distance, and player load – it had almost the poorest performance, and adding the player's playing position and sRPE to the imputation model did not improve performance of the logistic regression model to an acceptable level (Figure 6(a–c)). Under these conditions, Complete Case Analysis had the least bias (Figure 6(a–c)): 8% on average, and it was the only method with 100% coverage (Table 4).

Under MAR, Predicted Mean Matching was the only method with consistently acceptable bias, up to and including ≈50% missing data, when the total distance variable was the only variable missing (Figure 7(d–f)), and up to and including ≈25% missing when other GPS-variables were also missing and sRPE was in the imputation model (Figure 7(c)). Without sRPE, and with missing in all GPS variables, Predicted Mean Matching was not within acceptable bias (Figure 7(a,b)). The Mean Per Week was within acceptable bias in some cases (Figure 7(d–f)), but not in other cases (Figure 7(a–c)).

Discussion

This is the first systematic review of the training load and injury risk research field mapping the current practices of reporting and handling missing data. Only 34% of the 108 included studies reported whether the training load variable had any missing observations, 23% of the studies reported how the missing observations were handled, and 17% reported the amount of missing data.

Also, this study is the first attempt to determine the accuracy of discovering a relationship between training load and injury risk after handling missing data using different methods. All methods had acceptable accuracy when sRPE was missing ≤ 10% observations under the assumption of Missing Completely at Random (MCAR). However, only Multiple Imputation with Predicted Mean Matching accurately (≤ 3% bias) estimated the relationship through 50% missing, also under the assumption of Missing at Random (MAR). Being athlete reported data, sRPE is more likely to be MAR (Barnett et al. 2017).

For the GPS variable total distance, Multiple Imputation with Predicted Mean Matching was accurate up to and including 90% missing under MCAR and ≈50% under MAR, given that the other GPS variables were not missing too.

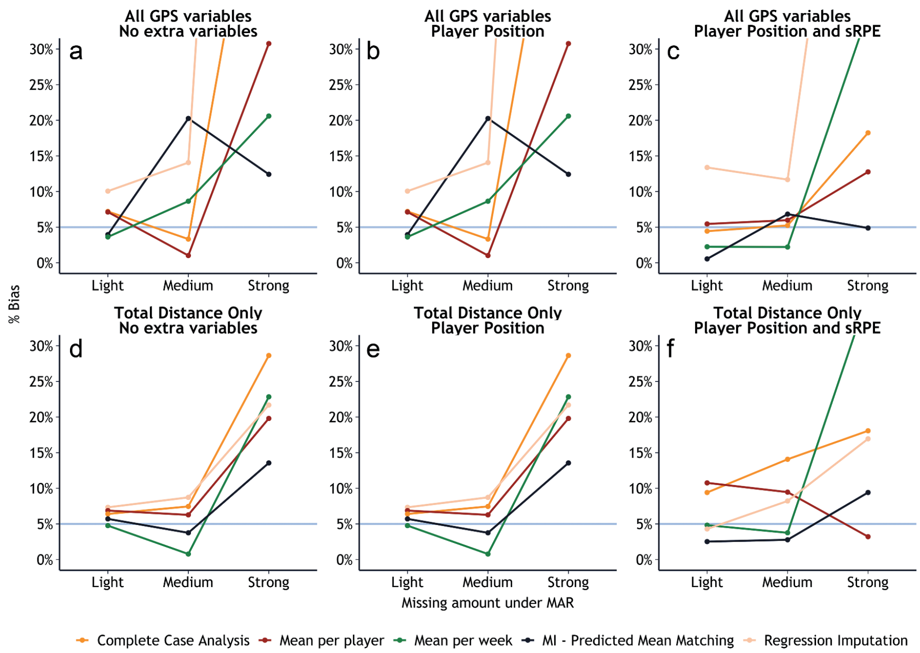

Figure 7. Performance of five different methods of imputing or deleting missing observations in training load measured by Global Positioning Systems (GPS) variables. Performance is measured by the Mean Percent Bias (% Bias). Varying levels of missing data under the assumption of Missing at Random (MAR) are displayed along the X-axis. Under MAR, the probability of missing is dependent on other variables. For (a–c), missing is introduced to all four GPS variables; (d–f) Missing is only introduced to the total distance variable. In addition, for (a, d) there were no extra variables in the imputation model, (b, e) the player position was in the imputation model, and (c, f) the session Rating of Perceived Exertion (sRPE) and the player position was in the imputation model. The best methods are closest to 0, and the light blue line indicates the maximum range of acceptable bias (0% to 5%). MI = Multiple imputation. Regression imputation and complete case analysis reached off the chart at > 90% bias in a–c. Logistic regression performed on the data without missing had a bias of 4%. Based on 1 900 simulations with monte carlo standard error < 0.00001.

Insufficient reporting of missing data in the training load and injury literature

In a recent review of articles on football topics, or involving football players, 11% of studies reported whether they had missing data (Borg et al. 2021). In our review of the training load and injury risk field, reporting practices were better (34%). Missing data were more thoroughly reported in clinical RCTs (72% Díaz-Ordaz et al. 2014) and in developmental psychology (57%, Jeličić et al. 2009), underlining that there is room for improvement in football research. Despite concerns raised by recent reviews (Windt et al. 2018; Griffin et al. 2020; Wang et al. 2021), our findings show no improvement in reporting missing data, varying between 30% and 50% in the last five years.

In a recent methodological review, 32% (N = 34) of studies reported how missing data was handled (Windt et al. 2018), and in a review on relative training load and injury (Andrade et al. 2020), 25% (N = 20) of studies reported it. Our findings suggest that reporting practices have not improved since these reviews were published. Many common methods of measuring relative training load (i.e. Acute:Chronic Workload Ratio) cannot be calculated in the presence of missing observations (Moussa et al. 2019), and therefore it is crucial to know how the missing data were handled.

Furthermore, without knowing the amount of missing data, it is impossible to deduce how they may have impacted the results. Only 17% of the studies reported both whether they had missing data and the amount of missing data, indicating severe reporting bias. Given that the mean amount of missing data reported by these studies was so low (7%, SD = 6) we suspect they represent a selection bias towards lower amounts of missing. In clinical cohort studies, where the amount of missing data was more certain (83% reported), it varied vastly, from 2% to 65% (Karahalios et al. 2012). This implies that underreporting of missing data is quite common in the field of training load and injury research.

Handling missing data in training load measures

Complete Case Analysis (also known as listwise deletion) retained the distribution of the observed sRPE and total distance covered data (supplementary file II Figures S3–S6). Despite this, it had high bias when attempting to detect a relationship between training load and injury risk. This shows that when imputed or complete case data appear to be like the observed data, it does not necessarily mean the method is valid.

In general, Mean Imputation had acceptable bias for ≤ 10% missing sRPE and ≤ 30% total distance covered under MCAR. How the mean was calculated for sRPE, player mean or weekly mean, was irrelevant; it did not alter the results in any meaningful way. For total distance covered, the two methods of calculation varied in performance when data were Missing at Random, and under this condition, the performance of Mean Imputation was rather unstable. We think Mean Imputation is only an acceptable approach when missing data amounts are small and, in situations where few variables associated with training load are available in the data.

Regression Imputation was outclassed by Multiple Imputation using Predicted Mean Matching (PMM). For sRPE, PMM had acceptable bias up to and including 50% under MCAR and MAR. For total distance covered, it reached unacceptable bias beyond 90% missing data and was the only method viable under MAR. These results reflect those in other simulations studies (Janssen et al. 2010; Knol et al. 2010). It had very poor performance, however, when the other GPS variables were missing. The accuracy did not improve with the player's playing position in the imputation model, although with the sRPE, it improved to a sufficient degree for < 25% missing under both MCAR and MAR. This finding highlights the importance of session-context information for GPS-measures, compared to player-context information. Also, this confirms the need for informative explanatory variables in the imputation model.

Multiple Imputation of PMM had lower bias than Single Imputation, but a higher standard error. This is caused by the between-imputation variation incorporated into the standard error as a surrogate measure of the uncertainty caused by the missing data itself (Sterne et al. 2009; Van Buuren 2018). The standard error, while higher, is, however, more representative of the true uncertainty.

Recommendations

We recommend future studies on training load and injury risk to report (1) whether they had missing data in training load, (2) the amount of missing data, and (3) how missing data were dealt with in the analyses.

Reporting the amount of missing data should include data that were consciously deleted, such as deletion of relative training load values in periods where athletes return from holidays (e.g. Hulin et al. 2016). We recommend using the checklist for the methods and results sections outlined by Borg et al. (2021) when reporting missing data.

The amount of missing data should be calculated as the number of missing observations divided by the number of potential observations. For sRPE, this is the number of missing RPE responses divided by the overall number of RPE prompts/questionnaires sent. For GPS measures, this is typically the number of missing daily GPS observations divided by the number of player days with an attached GPS device. We have provided an example of reporting this in the results section of this study (Page 5, line 315). In our study we were not able to calculate the amount of missing GPS data accurately, and we acknowledge that this may be the case for other studies as well. Therefore, we encourage transparency in how missing data is calculated.

To our knowledge, no other study has tested the order of events in imputing sRPE. In our analyses, imputing RPE and duration at the session level first, and then calculating sRPE, had the lowest bias (1.4%) and highest certainty. We recommend this approach in future studies.

Complete Case Analysis had the highest average confidence interval width in all scenarios, demonstrating loss of statistical power. In our systematic review, the median number of injuries was 85. We believe that study sample size and consequently statistical power can be improved by using imputation rather than applying deletion methods for missing data.

While Multiple Imputation is ideal under MAR, it does not solve the issue of Missing Not at Random (MNAR). Under MNAR, the probability of missing depends on information that is not available. Imagine a scenario where players with high training load are too busy to reply to the sRPE prompt. The sample will then be skewed towards lower sRPE values than the true population. Selection bias is inherent under MNAR and we advise caution in performing analyses on data where MNAR is assumed.

Multiple imputation methods can be used under MCAR and MAR. Other, more simpler methods, however, are appropriate under certain conditions only. We provide a guide for handling missing data of sRPE and GPS measures, respectively (Box 2).

Box 2. Recommended methods for handling missing data in training load.

Session rating of perceived exertion (sRPE)
sRPE is likely to be Missing at Random (MAR, Benson et al. 2021). However, this can be tested using Little's Missing Completely at Random (MCAR) test (Little 1988) and explored with visualizations (Borg et al. 2021).

- If low amounts of data are missing (approximately < 20%) under the assumption of MCAR, perform the following:
  - Complete Case Analysis if the remaining number of injuries will be of sufficient statistical power
  - Perform Mean Imputation if the number of injuries will not be of sufficient statistical power when using Complete Case Analysis
- If data are missing under the assumption of MAR, perform multiple imputation with Predicted Mean Matching

Global positioning systems
GPS data are more likely to be MCAR than sRPE, but the mechanism should be checked (Borg et al. 2021).

- If missing observations are intermittent in the GPS-variables, regardless of whether the data is assumed to be MCAR or MAR, perform multiple imputation with Predicted Mean Matching (PMM)
- If missing is consistent across all GPS-variables under the assumption of MCAR or MAR, and missing amount is low (approximately < 30%), Mean Imputation can be performed

A guide on how to perform Multiple Imputation with PMM, including pitfalls and solutions, is available at the primary author's GitHub site (reference link unavailable in anonymous version of thismanuscript).

Limitations

The systematic review was a limited search of the training load and injury risk field carried out by one author only. It was initially performed to provide a basis for methodological choices in the simulations, but the concerning results warranted an emphasis on how missing data should be reported. It may not accurately represent the entirety of the field of training load and injury risk research; endurance sports studies were especially underrepresented.

Due to anonymisation laws, the player's position on the team was not available in the uploaded dataset. We prioritized mimicking a realistic study scenario, and included this variable in multiple analyses, which are consequently irreproducible.

For the scope of this paper, we had to limit the number of methods compared. Specifically, k-nearest neighbour imputation and multiple imputation by random forest, which have shown successful performance in other simulation studies (Hasler and Tillé 2016; Chhabra et al. 2017), could be of interest in future studies. We also had to limit the number of conditions under which they were run. For instance, more levels of noise and sample sizes could be explored, as well the presence of non-linearity in the imputation model. PMM using a linear model may not be appropriate under such conditions (Morris et al. 2014; Bache-Mathiesen et al. 2021). Choosing to focus on only one GPS variable is also a limitation, and future studies should consider exploring high speed and sprint running distances, and more complex multivariate missing mechanisms between them (Schouten et al. 2018).

Finally, our simulations were based on a single dataset, and we had a limited number of explanatory variables available for use in our imputation models. Most notably, we think e.g. activity type, age, sex, and rehabilitation status would improve imputation models on training load and injury risk in future studies.

Conclusion

Our systematic review showed that in the field of training load and injury risk, reporting of missing data is insufficient. Future studies should report how missing data were handled in analyses. Multiple imputation with Predicted Mean Matching should be used to impute sRPE as well as GPS-variables and can be used in cases with large amounts of missing data, provided that the remaining data is representative of the population studied. We propose a guide for handling missing data in certain scenarios.

Acknowledgements

We thank Torstein Dalen-Lorentsen for providing access to the male Norwegian Premier League football data. We also thank Garth Theron for high quality programming of the Norwegian Premier League football database. This research had been impossible without the collaboration of coaches and players, and we thank the participants for their contribution.

Data availability

All data relevant to the study are included in the article or are available as supplementary files. The Norwegian Premier League football data were anonymised based on requirements of the Norwegian Data Protection Agency (Datatilsynet 2017), which required the removal of the variable describing the player's position on the football team. Any analysis which included this variable is therefore not reproducible. The review data on training load and injury risk studies are available as a machine-readable xlsx file, along with the anonymised Norwegian Premier League football data, and all statistical programming code, in a GitHub repository (Bache-Mathiesen 2021a).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethics Approval

The Norwegian Premier League football study was approved by the Ethical Review Board of the Norwegian School of Sport Sciences, and by the Norwegian Centre for Research Data (722773).

Twitter

@lena_kbm, @DocThorAndersen, @benclarsen, @FagerlandWang

Funding

The Oslo Sports Trauma Research Centre provided all funding for performing this study.

ORCID

L. K. Bache-Mathiesen http://orcid.org/0000-0002-9466-7003
Thor Einar Andersen http://orcid.org/0000-0003-4172-4518
Benjamin Clarsen http://orcid.org/0000-0003-3713-8938
Morten Wang Fagerland http://orcid.org/0000-0003-4535-5043

References

Albrecht J, Biese KM, Bell DR, Schaefer DA, Watson AM. 2020. Training load and injury among middle school–Aged athletes. J Athl Train. 55 (9):954–959. DOI:10.4085/1062-6050-435-19.

Andrade R, Wik EH, Rebelo-Marques A, Blanch P, Whiteley R, Espregueira-Mendes J, Gabbett TJ. 2020. 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 Med. 50(9) 1613–1635.

Bache-Mathiesen L 2021a. Missing data in training load (version 1.0.0). [Accessed 2021 OCt 14]. https://github.com/lenakba/missing-data-in-training-load

Bache-Mathiesen L 2021b. Training load and injury studies data file. GitHub: GitHub; [Accessed 2021 Oct 14]. https://github.com/lenakba/missing-data-in-training-load/blob/main/studies_missing_reporting.xlsx

Bache-Mathiesen LK 2021c. Performing multiple imputation in R. [Accessed 2021 Oct 14]. https://github.com/lenakba/missing-data-in-training-load/blob/main/guide-multiple-imputation.pdf

Bache-Mathiesen LK, Andersen TE, Dalen-Lorentsen T, Clarsen B, Fagerland MW. 2021. Not straightforward: modelling non-linearity in training load and injury research. BMJ Open Sport & Exercise Med. 7(3): e001119. DOI:10.1136/bmjsem-2021-001119.

Bahr R. 2016. Why screening tests to predict injury do not work—and probably never will ...: a critical review. Br J Sports Med. 50 (13):776–780. DOI:10.1136/bjsports-2016-096256.

Bahr R, Holme I. 2003. Risk factors for sports injuries—a methodological approach. Br J Sports Med. 37(5):384–392. DOI:10.1136/bjsm.37.5.384.

Bahr R, Krosshaug T. 2005. Understanding injury mechanisms: a key component of preventing injuries in sport. Br J Sports Med. 39(6):324–329. DOI:10.1136/bjsm.2005.018341.

Barnett AG, Mcelwee P, Nathan A, Burton NW, Turrell G. 2017. Identifying patterns of item missing survey data using latent groups: an observational study. BMJ Open. 7(10):e017284. DOI:10.1136/bmjopen-2017-017284.

Benson LC, Stilling C, Owoeye OB, Emery CA. 2021. Evaluating methods for imputing missing data from longitudinal monitoring of athlete workload. J Sports Sci Med. 20:187–195.

Borg DN, Nguyen R, Tierney NJ. 2021. Missing data: current practice in football research and recommendations for improvement. Sci Med Football. 1–6. DOI:10.1080/24733938.2021.1922739.

Borg G, Hassmén P, Lagerström M. 1987. Perceived exertion related to heart rate and blood lactate during arm and leg exercise. Eur J Appl Physiol Occup Physiol. 56(6):679–685. DOI:10.1007/BF00424810.

Bourdon PC, Cardinale M, Murray A, Gastin P, Kellmann M, Varley MC, Gabbett TJ, Coutts AJ, Burgess DJ, Gregson W. 2017. Monitoring athlete training loads: consensus statement. Int J Sports Physiol Perform. 12(s2): S2-161-S2–170. DOI:10.1123/IJSPP.2017-0208.

Boyd LJ, Ball K, Aughey RJ. 2011. The reliability of MinimaxX accelerometers for measuring physical activity in Australian football. Int J Sports Physiol Perform. 6(3):311–321. DOI:10.1123/ijspp.6.3.311.

Chhabra G, Vashisht V, Ranjan J. 2017. A comparison of multiple imputation methods for data with missing values. Indian J Sci Technol. 10(19):1–7. DOI:10.17485/ijst/2017/v10i19/110646.

Dalen-Lorentsen T, Andersen TE, Bjørneboe J, Vagle M, Martin KN, Kleppen M, Fagerland MW, Clarsen B. 2021. A cherry, ripe for picking: the relationship between the acute–chronic workload ratio and health problems. J Orthopaedic & Sports Physical Therapy. 51(4):162–173.

Datatilsynet. 2017. The anonymisation of personal data. *Norwegian Data Protection Agency*; [Accessed 2021 Oct 14]. https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/anonymisation

De Leeuw A-W, Van Der Zwaard S, Van Baar R, Knobbe A. 2021. Personalized machine learning approach to injury monitoring in elite volleyball players. Eur J Sport Sci. 1–14. DOI:10.1080/17461391.2021.1887369.

Demirtas H, Freels SA, Yucel RM. 2008. Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: a simulation assessment. J Stat Comput Simul. 78(1):69–84. DOI:10.1080/10629360600903866.

Díaz-Ordaz K, Kenward MG, Cohen A, Coleman CL, Eldridge S. 2014. Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines. Clin Trials. 11(5):590–600. DOI:10.1177/1740774514537136.

Dwyer DB, Gabbett TJ. 2012. Global positioning system data analysis: velocity ranges and a new definition of sprinting for field sport athletes. J Strength Condi Res. 26(3):818–824. DOI:10.1519/JSC.0b013e3182276555.

Eckard TG, Padua DA, Hearn DW, Pexa BS, Frank BS. 2018. The relationship between training load and injury in athletes: a systematic review. Sports Med. 48(8):1929–1961. DOI:10.1007/s40279-018-0951-z.

Enright K, Green M, Hay G, Malone JJ. 2019. Workload and injury in professional soccer players: role of injury tissue type and injury severity. Int J Sports Med. 41(02):89–97. DOI: 10.1055/a-0997-6741

Fanchini M, Rampinini E, Riggio M, Coutts AJ, Pecci C, Mccall A. 2018. Despite association, the acute: chronic work load ratio does not predict non-contact injury in elite footballers. Sci Med Football. 2(2):108–114. DOI:10.1080/24733938.2018.1429014.

Fernández-Cuevas I, Gómez-Carmona P, Sillero-Quintana M, Noya-Salces J, Arnaiz-Lastras J, Pastor-Barrón A. 2010. Economic costs estimation of soccer injuries in first and second spanish division professional teams. 15th Annual Congress of the European College of Sport Sciences ECSS; 23-26 june; Antalya, Turkey.

Foster C, Florhaug JA, Franklin J, Gottschall L, Hrovatin LA, Parker S, Doleshal P, Dodge C. 2001. A new approach to monitoring exercise training. J Strength and Condi Res. 15:10Y115.

Griffin A, Kenny IC, Comyns TM, Lyons M. 2020. The association between the acute: chronic workload ratio and injury and its application in team sports: a systematic review. Sports Med. 50(3):561–580.

Hägglund M, Waldén M, Magnusson H, Kristenson K, Bengtsson H, Ekstrand J. 2013. Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. Br J Sports Med. 47(12):738–742. DOI:10.1136/bjsports-2013-092215.

Hasler C, Tillé Y. 2016. Balanced k-nearest neighbour imputation. Stat. 50 (6):1310–1331. DOI:10.1080/02331888.2016.1230615.

Hulin BT, Gabbett TJ, Lawson DW, Caputi P, Sampson JA. 2016. The acute: chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. Br J Sports Med. 50 (4):231–236. DOI:10.1136/bjsports-2015-094817.

Janssen KJ, Donders ART, Harrell FE JR, Vergouwe Y, Chen Q, Grobbee DE, Moons KG. 2010. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol. 63(7):721–727. DOI:10.1016/j.jclinepi.2009.12.008.

Jeličić H, Phelps E, Lerner RM. 2009. Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology. Dev Psychol. 45(4):1195. DOI:10.1037/a0015665.

Karahalios A, Baglietto L, Carlin JB, English DR, Simpson JA. 2012. A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. BMC Med Res Methodol. 12 (1):96. DOI:10.1186/1471-2288-12-96.

Knol MJ, Janssen KJ, Donders ART, Egberts AC, Heerdink ER, Grobbee DE, Moons KG, Geerlings MI. 2010. Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. J Clin Epidemiol. 63(7):728–736. DOI:10.1016/j.jclinepi.2009.08.028.

Lang T 2004. Twenty statistical errors even you can find in biomedical research articles. Citeseer.

Lang T. 2005. Common statistical errors even you can find. part 6: errors in research designs. Part. 6:112–115.

Little RJ. 1988. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc. 83(404):1198–1202. DOI:10.1080/01621459.1988.10478722.

Lolli L, Bahr R, Weston M, Whiteley R, Tabben M, Bonanno D, Gregson W, Chamari K, Di Salvo V, Van Dyk N. 2020. No association between perceived exertion and session duration with hamstring injury occurrence in professional football. Scand J Med Sci Sports. 30(3):523–530. DOI:10.1111/sms.13591.

Malone JJ, Lovell R, Varley MC, Coutts AJ. 2017. Unpacking the black box: applications and considerations for using GPS devices in sport. Int J Sports Physiol Perform. 12(s2):S2-18-S2–26. DOI:10.1123/ijspp.2016-0236.

Mansournia MA, Collins GS, Nielsen RO, Nazemipour M, Jewell NP, Altman DG, Campbell MJ. 2021. A checklist for statistical assessment of medical papers (the CHAMP statement): explanation and elaboration. Br J Sports Med. 55(18):1009-1017. DOI:10.1136/bjsports-2020–103652.

Maupin D, Schram B, Canetti E, Orr R. 2020. The relationship between acute: chronic workload ratios and injury risk in sports: a systematic review. Open Access J Sports Med. 11:51. doi:10.2147/OAJSM.S231405.

Mccall A, Dupont G, Ekstrand J. 2018. Internal workload and non-contact injury: a one-season study of five teams from the UEFA elite club injury study. Br J Sports Med. 52(23):1517–1522. DOI:10.1136/bjsports-2017-098473.

Moreno-Pérez V, Prieto J, Del Coso J, Lidó-Micó JE, Fragoso M, Penalva FJ, Reid M, Pluim BM. 2020. Association of acute and chronic workloads with injury risk in high-performance junior tennis players. Eur J Sport Sci. 21 (8):1215–1223.

Morris TP, White IR, Crowther MJ. 2019. Using simulation studies to evaluate statistical methods. Stat Med. 38(11):2074–2102. DOI:10.1002/sim.8086.

Morris TP, White IR, Royston P. 2014. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Med Res Methodol. 14(1):75. DOI:10.1186/1471-2288-14-75.

Moussa I, Leroy A, Sauliere G, Schipman J, Toussaint J-F, Sedeaud A. 2019. Robust exponential decreasing index (REDI): adaptive and robust method for computing cumulated workload. BMJ Open Sport & Exercise Med. 5(1):e000573. DOI:10.1136/bmjsem-2019-000573

Musil CM, Warner CB, Yobas PK, Jones SL. 2002. A comparison of imputation techniques for handling missing data. West J Nurs Res. 24(7):815–829. DOI:10.1177/019394502762477004.

Nielsen RO, Bertelsen ML, Ramskov D, Møller M, Hulme A, Theisen D, Finch CF, Fortington LV, Mansournia MA, Parner ET. 2019. Time-to-event analysis for sports injury research part 2: time-varying outcomes. Br J Sports Med. 53(1):70–78. DOI:10.1136/bjsports-2018-100000.

Nielsen RO, Shrier I, Casals M, Nettel-Aguirre A, Møller M, Bolling C, Bittencourt NF, Clarsen B, Wedderkopp N, Soligard T. 2020. Statement on methods in sport injury research from the first METHODS MATTER meeting, copenhagen, 2019. J Orthopaedic & Sports Physical Therapy. 50(5):226–233. DOI:10.2519/jospt.2020.9876.

Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE JR, Moons KG, Collins GS. 2019. Minimum sample size for developing a multivariable prediction model: PART II-binary and time-to-event outcomes. Stat Med. 38 (7):1276–1296. DOI:10.1002/sim.7992.

Rønneberg KL. 2020. 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, Oslo, Norway.

Schouten RM, Lugtig P, Vink G. 2018. Generating missing values for simulation purposes: a multivariate amputation procedure. J Stat Comput Simul. 88(15):2909–2930. DOI:10.1080/00949655.2018.1491577.

Schouten RM, Vink G. 2018. The dance of the mechanisms: how observed information influences the validity of missingness assumptions. Sociol Methods Res. 50(3):1243–1258. DOI: 10.1177/0049124118799376.

Shmueli G. 2010. To explain or to predict? Stat Sci. 25(3):289–310. DOI:10.1214/10-STS330.

Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 338(1):b2393. DOI:10.1136/bmj.b2393.

Theisen D, Frisch A, Malisoux L, Urhausen A, Croisier J-L, Seil R. 2013. Injury risk is different in team and individual youth sport. J Sci Med Sport. 16 (3):200–204. DOI:10.1016/j.jsams.2012.07.007.

Theron GF. 2020. The use of data mining for predicting injuries in professional football players. University of Oslo, Oslo, Norway.

Udby CL, Impellizzeri FM, Lind M, Nielsen RØ. 2020. How has workload been defined and how many workload-related exposures to injury are included in published sports injury articles? A scoping review. J Orthopaedic & Sports Physical Therapy. 50(10):538–548. DOI:10.2519/jospt.2020.9766.

Van Buuren S. 2018. Flexible imputation of missing data. Chapman and Hall, CRC Press, Boca Ration, Florida, USA.

Van Smeden M, De Groot JA, Moons KG, Collins GS, Altman DG, Eijkemans MJ, Reitsma JB. 2016. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Med Res Methodol. 16(1):1–12. DOI:10.1186/s12874-016-0267-3.

Vink G 2016. Towards a standardized evaluation of multiple imputation routines. [Accessed 2021 Oct 14]. https://www.gerkovink.com/docs/Standardized%20evaluation.pdf

Von Hippel PT. 2009. 8. How to impute interactions, squares, and other transformed variables. Sociol Methodol. 39(1):265–291. DOI:10.1111/j.1467-9531.2009.01215.x.

Wang A, Healy J, Hyett N, Berthelot G, Okholm Kryger K. 2021. A systematic review on methodological variation in acute: chronic workload research in elite male football players. Sci Med Football. 5(1):18–34. DOI:10.1080/24733938.2020.1765007.

White IR, Carlin JB. 2010. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat Med. 29(28):2920–2931. DOI:10.1002/sim.3944.

White IR, Royston P, Wood AM. 2011. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 30(4):377–399. DOI:10.1002/sim.4067.

Windt J, Ardern CL, Gabbett TJ, Khan KM, Cook CE, Sporer BC, Zumbo BD. 2018. Getting the most out of intensive longitudinal data: a methodological review of workload–injury studies. BMJ Open. 8(10):e022626. DOI:10.1136/bmjopen-2018-022626.

Windt J, Gabbett TJ. 2017. How do training and competition workloads relate to injury? The workload—injury aetiology model. Br J Sports Med. 51 (5):428–435. DOI:10.1136/bjsports-2016-096040.

Supplementary Methods

Ethics

The study was approved by the Ethical Review Board of the Norwegian School of Sport Sciences, and the Norwegian Centre for Research Data (722773). Ethical principles were followed in accordance with the Declaration of Helsinki (Malik & Foster, 2016), 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). All participants provided informed written consent. Data were anonymised according to guidelines outlined by The Norwegian Data Protection Authority (Datatilsynet, 2017). The datasets cannot be joined.

Training load definition

Session Rating of Perceived Exertion

In 17 of 4 725 (0.4%) instances where players reported participating in training, but with a duration of 0 minutes, RPE was assumed to be 0 (no participation). In 73 (2%) instances where players reported participating in training with a duration above 0 minutes, yet RPE was reported at 0 (meaning no participation), RPE was assumed to be 1.

Global positioning systems

GPS data were attained per second and transferred to a database with an automatic cleaning procedure, altogether 14 611 560 total distance values (Theron, 2020). Errors were handled as follows:

- Total distance > 30 meters per second were removed (79 observations).
- Total distance > 15 meters & <= 30 meters per second were considered overestimated outliers and set to 15, following guidelines in Harrell Jr (2017) (348 observations).

Daily total distance was calculated. Of 2 984 daily total distance values, 92 were less than 100 meters per day and considered missing.

Simulations

Comparison of Imputation Methods

All missing was removed from all variables in the datasets. The final datasets had 4 725 sRPE values and 2 292 total distance values of Gaussian distributions (supplemental file II Figure S1). Simulated injuries were added to the sRPE dataset with a predefined, relationship with sRPE, using a logistic regression function:

$$logistic(x) = \frac{1}{1 + \exp(-x)}$$

$$Prob(Injury) = logistic(-2 + 0.003 * sRPE)$$

Likewise, for total distance:

$$Prob(Injury) = logistic(-2 + 0.0003 * Total\ Distance)$$

Logistic regression was run with injury as the response variable and training load as the explanatory variable to determine base performance when no data were missing.

Imputation strategy for a derived variable

It is unknown how a derived variable such as sRPE should be imputed (Benson et al., 2021; Van Buuren, 2018). Simulations were performed to compare four strategies of imputing sRPE:

- Impute duration and RPE without sRPE in the dataset, and calculate sRPE after imputation (Impute, then transform, Von Hippel, 2009). In this method, the product (sRPE) is not available to inform the imputation model. However, it may reduce collinearity issues.
- Calculate sRPE, and impute duration, RPE and sRPE as regular variables (Transform, then impute, Von Hippel, 2009; White et al., 2011). Here, sRPE is present to inform the imputation model.
- Calculate sRPE and impute, but add the relationship between RPE, duration and sRPE in the imputation model, thereby transforming on-the-fly within the imputation algorithm (Passive imputation, Van Buuren, 2018). This may be an improvement over using them merely as explanatory variables.
- We hypothesized that other variables, such as whether the activity is a match, may be enough information to impute sRPE. Therefore, we also tried removing duration and RPE from the dataset and relying only on other variables for information when imputing sRPE (Impute product without factors). Under this scenario, no issues stemming from the strong correlation between intensity, duration and sRPE are present, but intensity and duration are not available to inform the imputation model. This may be reasonable for studies which only have access to the product, sRPE.

Fake injuries were added to the dataset with a predefined relationship with sRPE:

$$Prob(Injury) = logistic(-2 + 0.003 * sRPE)$$

We tested the ability of a logistic regression model to discover this relationship in a dataset with 25% missing RPE and duration values. For each of the four methods listed above, the missing observations were imputed with predicted mean matching with 1 900 permutations. The most accurate method determined in these simulations were used in Step 2 below.

Step 1 Add missing

We hypothesized that missing observations in total distance and other GPS variables are likely to be Missing Completely at Random (MCAR). Under MCAR, all observations have an equal probability of missing – the probability is not dependent on other factors. For instance, we can imagine that technical errors in a GPS device can happen at random, and are not dependent on the characteristics of the athlete or the performed activity.

From the sRPE dataset and total distance dataset, eleven datasets were created with amounts of missing sampled at random under the assumption of MCAR: 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. This follows recommendations to cover a range of percentages of missing data as opposed to a select few (Vink, 2016).

For RPE – an athlete-reported parameter – we theorized that it is more likely that the probability of missing data depends on characteristics of the player and also of the day of activity, which was the case in Benson et al. (2021). This assumption is known as Missing at Random (MAR, Janssen et al., 2010). As a hypothetical example, players may be busier on match days and forget to report RPE. In such a scenario, whether a day is a match can predict the probability of missing RPE data. Missing GPS data may also be MAR if, for instance, the devices were worn in different locations and environmental obstruction was present in some locations and not in others (Malone et al., 2017).

Simulation studies have shown that weak correlations under a MAR mechanism may mimic MCAR to such a degree that methods which typically underperform under MAR, may perform to a sufficient degree (Schouten & Vink, 2018). To ensure we can identify this phenomenon, three datasets were created from the sRPE dataset and total distance dataset with different levels of missing under the assumption of MAR, with increasingly stronger correlations between variables and the probability of missing in training load. Fake ages were drawn at random from 18 to 30 years and added to the dataset. Fake sex was randomly sampled from female 0 and male 1. A recovery day was defined as the day after a match or two days after a match (M+1 or M+2), and coded as 1 for recovery day, 0 for training day. Match was coded as 1 for match, 0 for no match. Missing was then added with the following probability functions:

Light MAR (≈25% missing):

$$Prob(Missing) = logistic(-2 + 0.03 * Age + \ 0.02 * Sex + 0.3 * Recovery\ day)$$

Medium MAR (≈50% missing):

$$Prob(Missing) = logistic(-2 + 0.08 * Age + \ 0.04 * Sex + 0.8 * Recovery\ day)$$

Strong MAR (≈80% missing):

$$Prob(Missing) =$$

$$logistic(-2 + 0.13 * Age + \ 0.1 * Sex + 1.8 * Recovery\ day + 1.8 * Match)$$

Thus, for the two datasets, sRPE and total distance, 11 MCAR datasets and 3 MAR datasets were generated.

Step 2 Impute or delete missing

The missing observations in the 14 sRPE and 14 total distance datasets created in Step 1 were imputed or deleted with five different methods, respectively.

Complete Case Analysis

The simplest form of handling missing data is to delete the rows with missing observations. The analyses are then run only on the complete cases. The statistical software packages SPSS, R and Stata use Complete Case Analysis (listwise deletion) as the default method if nothing else is specified (IBM, 2020; Kabacoff, 2011; UCLA, 2021). Besides its simplicity, Complete Case Analysis also has the advantage that it cannot introduce unrealistic or impossible values. However, it reduces the sample size and may introduce selection bias (White & Carlin, 2010).

Mean Imputation

Replacing the missing value with the mean of the parameter is known as Mean Imputation (Barzi & Woodward, 2004). This was the most popular imputation method in the collected training load studies (31%, Table 1). However, there was a wide discrepancy in how the articles calculated the mean (Colby et al., 2014; O'Keeffe et al., 2019; Sampson et al., 2019; Windt et al., 2017), a concern raised in Wang et al. (2021). Benson et al. (2021) defined two overarching categories of means: individual context means and session context means. We compared two methods of Mean Imputation: the player mean, representing an individual context, and the weekly mean, representing a session context.

Table 1. The methods used to handle missing observations in the training load variable as reported by 36[1] studies in the field of training load and injury risk research.

Missing Data Method	N studies	% of studies
Unclear[2]	12	33%
Mean Imputation	11	31%
Complete Case Analysis	8	22%
Median Imputation	2	6%
Multiple Imputation	2	6%
Regression Imputation	1	3%

[1] Of 108 eligible studies, 37 (34%) reported whether they had missing data in the training load variable. One of the 37 studies had no missing data and was removed from this analysis.
[2] Methods considered “Unclear” were cases where authors reported having missing data, but the method used could not be determined.

Mean Imputation has the advantage of simplicity; it is easy to understand and implement (Benson et al., 2021). The downside of Mean Imputation is that it may reduce the variability of the dataset, and skew distributions (Barzi & Woodward, 2004). In addition, its performance depends on how the means are calculated. One potential pitfall is using means that may not be representative of the true observation (Bowen et al., 2020).

Multiple Imputation with Predicted Mean Matching

In Multiple Imputation, other variables in the dataset are used to predict the imputations for the missing observations (Van Buuren, 2018). Each missing observation is predicted multiple times, creating m number of imputed datasets. The regression model of interest is then run on each of the m datasets. The estimates from each model are averaged using Rubin's rules, which calculate standard errors that account for between-imputation variation and the level of uncertainty that stems from the missing data (Sterne et al., 2009; Van Buuren, 2018). See Figure 1 for an illustration of the Multiple Imputation process.

The observations used to impute the missing observations may be predicted with different methods. We chose Predicted Mean Matching (PMM). In PMM, all values for the target variable (training load), whether they are missing or not, are predicted using the other variables in the dataset. In this study, a linear regression model was used (Van Buuren, 2018). PMM forms a set of candidate donors from all complete cases with predicted values closest to the predicted value for the missing observation. From these candidates, one donor is randomly drawn. The observed value of that donor is used to impute the missing observation. Here, the number of candidate donors to draw from was 5 (Van Buuren, 2018).

PMM has the advantage that imputations outside the observed data range cannot occur (such as negative training loads), and it is less vulnerable to model misspecification than other methods (Little & Rubin, 2019). A disadvantage is that it may use the same donor multiple times, and in small sample sizes, cause superficially low variability (Van Buuren, 2018).

We ran Multiple Imputation with five imputed datasets (Van Buuren, 2018). A separate analysis was performed to compare multiple and single imputation of PMM.

Figure 1. Illustration of the modeling process in the framework of multiple imputation. In the first step, the variables available in the dataset are used to predict m number of potential imputations for the missing observations. A dataset is created for each of the m sets of predictions. The number of imputed datasets, five, is recommended in most cases (Van Buuren, 2018, section 2.8). The main model of interest is then run on each of the 5 datasets. The estimates from each model are averaged using Rubin's rules, which calculate standard errors that account for between-imputation variation and the level of uncertainty that stems from the missing data (Sterne et al., 2009; Van Buuren, 2018).

Regression Imputation

One training load and injury study reported using Regression Imputation (Table 1, Esmaeili et al., 2018). In Regression Imputation, the missing observation is replaced with the predicted value from a linear regression model. An advantage of Regression Imputation is that the true variability in the data may be more preserved than using Mean Imputation methods. However, the linear regression model cannot predict outliers, which may cause the distribution of imputed values to become unrealistically smooth.

Imputation model

We performed Predicted Mean Matching and Regression Imputation with the default imputation model from the R package mice (Buuren, 2011). It conditions the predictions on all other variables available in the dataset. These were: player ID (anonymised); date of activity; match (yes/no); micro-cycle-day (M, M-1, M-2, M-3, M-4, M+2, M+1). The simulated response variable, injury, was also used to predict imputed values (Moons et al., 2006; Sterne et al., 2009). If missing is under the assumption of MAR, variables that predict missing should not be included in the imputation model (Van Buuren, 2018). We therefore did not include the simulated sex and age, nor recovery day. In the total distance dataset, the other GPS variables were also in the imputation model: high-speed running distance, sprint distance, and player load.

Step 3 Fitting models on imputed data

Logistic regression models were run with training load as the independent variable, and the simulated injuries as the outcome variable, on each of the 14 sRPE and 14 total distance datasets imputed in Step 2. For Multiple Imputation, a logistic regression model was fit on each of the 5 generated datasets, and the results were pooled using Ruben's rules (Van Buuren, 2018).

In summary, the three steps of the simulations were:

1. Add missing drawn under MCAR and MAR from the original dataset
2. Impute or delete missing data using 5 different methods
3. Fit logistic regression models with injury as the outcome and training load as the explanatory variable on the imputed or missing-omitted data

Performance Measures

The following performance measures are recommended in Van Buuren (2018) and were used to compare the performance of the different imputation methods:

- The Raw Bias (RB) was the absolute difference between the estimated coefficient and the observed coefficient $(\hat{\theta} - \theta)$. Let $\hat{\theta}$ denote the estimated coefficient from running a logistic regression model on imputed data. The formula for raw bias in this study was for sRPE, $\hat{\theta} - 0.003$, and for total distance, $\hat{\theta} - 0.0003$.
- Percent bias (PB). The upper limit for acceptable performance was ±5% (Demirtas et al., 2008).
- Root-Mean-Squared Error (RMSE). A compromise between bias and variance that evaluates the estimated coefficient $\hat{\theta}$ on both accuracy and precision. If all methods have acceptable bias, they may be distinguished by RMSE.
- Coverage: the proportion of 95% confidence intervals that contained the true value.
- Average width (AW) of the 95% confidence intervals, which is an indicator of statistical efficiency. If all methods have a coverage > 95%, they may be distinguished by average width.

Using formulas listen in Morris et al. (2019), accepting a Monte Carlo Standard Error of no more than 0.5, the number of permutations 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 permutations needed for an accurate estimate of bias was calculated by:

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

Where s^2 is the sample variance (Morris et al., 2019). For an estimation of variance, a pilot of 100 permutations were run. The variance of the bias was < 0.00001; the number of permutations needed to achieve the target MCSE was < 100. Since coverage required more permutations to achieve target MCSE, simulation steps 1–3 outlined above were repeated 1 900 times. The mean of each performance measure was calculated across the 1 900 simulations.

The percent bias per method was visualized for each scenario of missing. In addition, a visualization of imputed vs. observed data was created for 50% missing under MCAR and ≈80% missing under MAR, to see if imputation methods managed to retain the properties of the observed data.

Imputation with extra variables available

The simulations were repeated to test whether the results changed with the inclusion or exclusion of the player's playing position in the imputation model.

In addition, for the total distance analyses, we considered that some studies collect daily sRPE alongside GPS measures. Further, we considered that in many cases, if the total distance GPS measure is missing, it is likely that the other GPS measures are also missing. Following recommendations to include multivariate missing, and consider more realistic scenarios (Schouten et al., 2018), we performed the simulations outlined previously (steps 1 to 3) under six scenarios for total distance:

1. Missing was added to total distance only, and no extra variables were in the dataset.
2. Missing was added to total distance only, and the player's playing position was among the variables in the dataset.
3. Missing was added to total distance only, and both the player's playing position and the sRPE was among the variables in the dataset.
4. Missing was added to all the GPS variables, and no extra variables were in the dataset.
5. Missing was added to all the GPS variables, and the player's playing position was among the variables in the dataset.
6. Missing was added to all the GPS variables, and both the player's playing position and the sRPE was among the variables in the dataset.

Data tools

All statistical analyses and simulations were performed using R version 4.1.0 (R Core Team, 2021) with RStudio version 1.4.1717. The MICE package was used for Multiple Imputation (Buuren, 2011), the visdat package for visualizing missing data (Tierney, 2017), and chron for manipulating time data (James, 2020). The simulations were run on a computer with an Intel(R) Core(TM) i6-8265U 1.6GHz CPU, and with 8 GB RAM. A GitHub repository is available with all R code and the data used in the simulations (Bache-Mathiesen, 2021).

References

Bache-Mathiesen, L. K. (2021). *Missing data in training load: R code repository*. In (Version 1.0) [R code repository]. GitHub. https://github.com/lenakba/missing-data-in-training-load

Barzi, F., & Woodward, M. (2004). Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *American journal of epidemiology*, *160*(1), 34-45. https://doi.org/10.1093/aje/kwh175

Benson, L. C., Stilling, C., Owoeye, O. B., & Emery, C. A. (2021). Evaluating Methods for Imputing Missing Data from Longitudinal Monitoring of Athlete Workload. *Journal of Sports Science and Medicine*, *20*(2), 187-195. https://doi.org/10.52082/jssm.2021.188

Bowen, L., Gross, A. S., Gimpel, M., Bruce-Low, S., & Li, F.-X. (2020). Spikes in acute: chronic workload ratio (ACWR) associated with a 5–7 times greater injury rate in English Premier League football players: a comprehensive 3-year study. *British journal of sports medicine*, *54*, 731-738. https://doi.org/10.1136/bjsports-2018-099422

Buuren, S. v. G.-O., Karin. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1-67. https://doi.org/10.18637/jss.v045.i03

Colby, M. J., Dawson, B., Heasman, J., Rogalski, B., & Gabbett, T. J. (2014). Accelerometer and GPS-derived running loads and injury risk in elite Australian footballers. *The Journal of Strength & Conditioning Research*, *28*(8), 2244-2252.

Datatilsynet. (2017). *The anonymisation of personal data*. Norwegian Data Protection Agency. Retrieved 2022-11-07 from https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/anonymisation/

Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: a simulation assessment. *Journal of Statistical Computation and Simulation*, *78*(1), 69-84. https://doi.org/10.1080/10629360600903866

Esmaeili, A., Hopkins, W. G., Stewart, A. M., Elias, G. P., Lazarus, B. H., & Aughey, R. J. (2018). The individual and combined effects of multiple factors on the risk of soft tissue non-contact injuries in elite team sport athletes. *Frontiers in physiology*, *9*. https://doi.org/10.3389/fphys.2018.01280

Harrell Jr, F. E. (2017). Regression modeling strategies. *BIOS*, *330*, 2018. https://link.springer.com/book/10.1007/978-3-319-19425-7

IBM. (2020, 2020-04-16). *Pairwise vs. Listwise deletion: What are they and when should I use them?* IBM Support. Retrieved 2022-11-07 from https://www.ibm.com/support/pages/pairwise-vs-listwise-deletion-what-are-they-and-when-should-i-use-them

James, D. H., Kurt. (2020). *chron: Chronological Objects which Can Handle Dates and Times*. In (Version version 2.3-56) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=chron

Janssen, K. J., Donders, A. R. T., Harrell Jr, F. E., Vergouwe, Y., Chen, Q., Grobbee, D. E., & Moons, K. G. (2010). Missing covariate data in medical research: to impute is better than to ignore. *Journal of clinical epidemiology*, *63*(7), 721-727. https://doi.org/10.1016/j.jclinepi.2009.12.008

Kabacoff, R. (2011). *R in Action*. Manning publications Shelter Island, NY, USA. https://www.manning.com/books/r-in-action-second-edition

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Malik, A. Y., & Foster, C. (2016). The revised Declaration of Helsinki: cosmetic or real change? *Journal of the Royal Society of Medicine*, *109*(5), 184-189.

Malone, J. J., Lovell, R., Varley, M. C., & Coutts, A. J. (2017). Unpacking the black box: applications and considerations for using GPS devices in sport. *International journal of sports physiology and performance*, *12*(s2), S2-18-S12-26. https://doi.org/10.1123/ijspp.2016-0236

Moons, K. G., Donders, R. A., Stijnen, T., & Harrell Jr, F. E. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of clinical epidemiology*, *59*(10), 1092-1101.

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in medicine*, *38*(11), 2074-2102. https://doi.org/10.1002/sim.8086

O'Keeffe, S., O'Connor, S., & Ní Chéilleachair, N. (2019). Are internal load measures associated with injuries in male adolescent Gaelic football players? *European Journal of Sport Science*, 1-12.

R Core Team. (2021). *R: A language and environment for statistical computing.* In [Computer program]. Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Sampson, J. A., Murray, A., Williams, S., Sullivan, A., & Fullagar, H. H. (2019). Subjective Wellness, Acute: Chronic Workloads, and Injury Risk in College Football. *Journal of strength and conditioning research*, *33*(12), 3367-3373. https://doi.org/10.1519/JSC.0000000000003000

Schouten, R. M., Lugtig, P., & Vink, G. (2018). Generating missing values for simulation purposes: a multivariate amputation procedure. *Journal of Statistical Computation and Simulation*, *88*(15), 2909-2930.

Schouten, R. M., & Vink, G. (2018). The dance of the mechanisms: How observed information influences the validity of missingness assumptions. *Sociological Methods & Research*, 0049124118799376.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, *338*, b2393. https://doi.org/10.1136/bmj.b2393

Theron, G. F. (2020). *The use of Data Mining for Predicting Injuries in Professional Football Players* [Master thesis, University of Oslo]. http://urn.nb.no/URN:NBN:no-81831

Tierney, N. (2017). visdat: Visualising whole data frames. *Journal of Open Source Software*, *2*(16), 355. https://doi.org/10.21105/joss.00355

UCLA. (2021). *Missing values | STATA learning modules*. UCLA: Statistical Consulting Group. Retrieved 2021-03-24 from https://stats.idre.ucla.edu/stata/modules/missing-values/

Van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman and Hall/CRC. https://stefvanbuuren.name/fimd/

Vink, G. (2016). *Towards a standardized evaluation of multiple imputation routines*. Retrieved 2022-11-07 from https://www.gerkovink.com/published/%5Bworking%5D%20Towards%20a%20standardized%20evaluation%20of%20multiple%20imputation%20routines.pdf

Von Hippel, P. T. (2009). 8. How to impute interactions, squares, and other transformed variables. *Sociological methodology*, *39*(1), 265-291. https://doi.org/10.1111/j.1467-9531.2009.01215.x

Wang, A., Healy, J., Hyett, N., Berthelot, G., & Okholm Kryger, K. (2021). A systematic review on methodological variation in acute: chronic workload research in elite male football players. *Science and Medicine in Football*, *5*(1), 18-34. https://doi.org/10.1080/24733938.2020.1765007

White, I. R., & Carlin, J. B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in medicine*, *29*(28), 2920-2931. https://doi.org/10.1002/sim.3944

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*, *30*(4), 377-399. https://doi.org/10.1002/sim.4067

Windt, J., Gabbett, T. J., Ferris, D., & Khan, K. M. (2017). Training load--injury paradox: is greater preseason participation associated with lower in-season injury risk in elite rugby league players? *British journal of sports medicine*, *51*(8), 645-650. https://doi.org/10.1136/bjsports-2016-095973

Supplementary Results

Tables

Table S1. A comparison of Multiple (MI) and Single Imputation (SI) of missing observations in total distance using Predicted Mean Matching.

		% Bias		Standard Error	
		MI	SI	MI	SI
Mean		3.5	4.1	0.0000218	0.0000206
Missing %	5%	4	3.2	0.0000205	0.0000204
	10%	3.6	4.6	0.0000207	0.0000207
	20%	5.6	4.4	0.0000208	0.0000206
	30%	1.8	1.3	0.0000209	0.0000203
	40%	2.1	0.6	0.0000209	0.0000201
	50%	5.7	6.1	0.0000211	0.0000208
	60%	3.1	5.7	0.0000231	0.0000209
	70%	3.5	3.2	0.0000232	0.0000209
	80%	2.7	3.9	0.0000241	0.0000200
	90%	3.2	8	0.0000229	0.0000209

Table S2. The absolute raw and percent bias, root-mean-squared-error, coverage and average width of 95% confidence intervals, for four methods of imputing sRPE before running logistic regression modelling a predefined, known relationship between sRPE and injury probability. Number of simulations = 1 900.

Imputation Method	Raw Bias	% Bias	RMSE	MCSE for RMSE	CR	MCSE for CR	AW
Impute, then transform	0.0000427	1.4%	0.0000124	<0.0001	100%	0.5	0.000745
Transform, then impute	0.0000791	2.6%	0.0001100	<0.0001	100%	0.5	0.000943
Passive imputation	0.0000759	2.5%	0.0000894	<0.0001	100%	0.5	0.000713
Impute product without factors	0.0000525	1.8%	0.0000599	<0.0001	100%	0.5	0.000895

Abbreviations: AW, Average width; CR, coverage; MCSE, Monte Carlo Standard Error; sRPE, session Rating of Perceived Exertion; RMSE, Root-Mean-Squared-Error

Figures

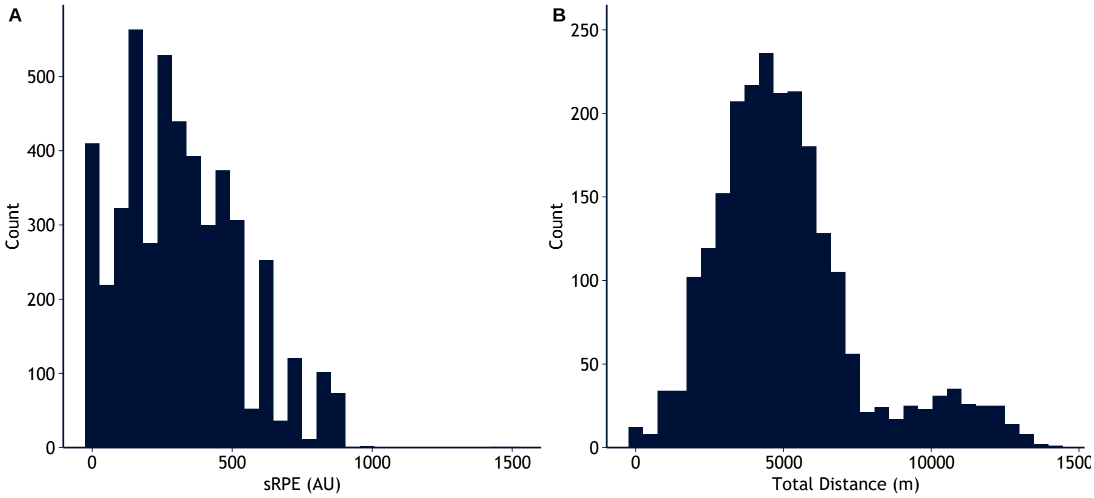

Figure S1. Distribution of training load variables in a male Norwegian Premier League football dataset. The variables were used in a simulation to compare the performance of imputing missing data in training load. For (A) session Rating of Perceived Exertion (sRPE) measured in Arbitrary Units (AU), and (B) total distance measured in meters (m), based on 4 725 sRPE values and 2 292 total distance values of 39 male professional football players.

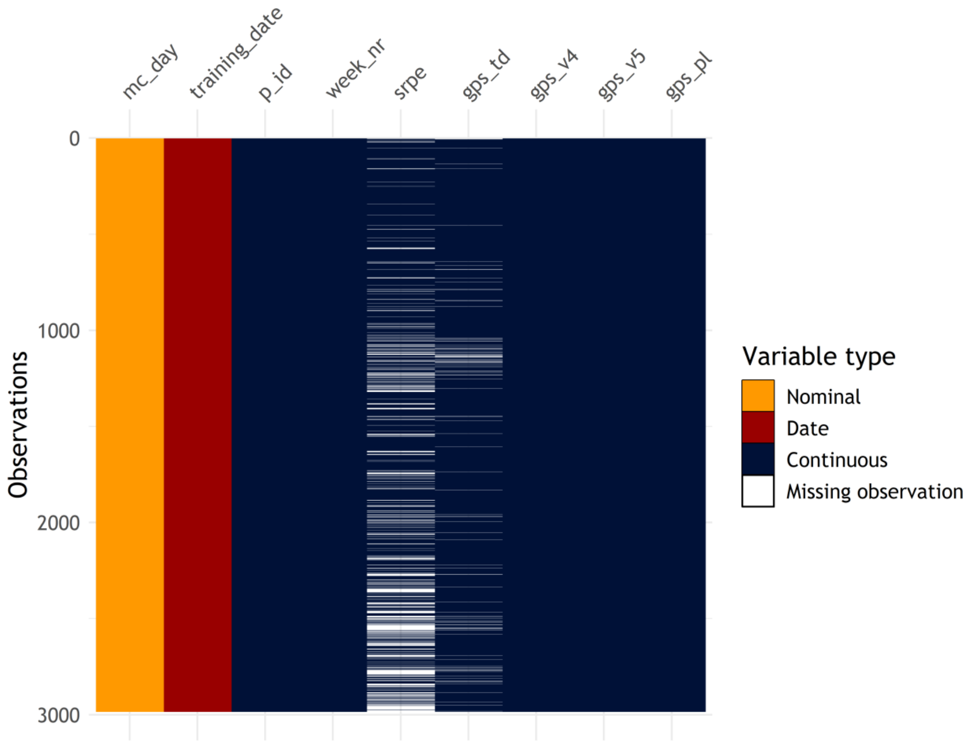

Figure S2. Distribution of missing observations in the Norwegian Premier League football dataset. The data is arranged from the beginning of the study period (top) to the end of the study (bottom). Variables were micro-cycle-day (mc_day), date of activity (training_date), player identifier (p_id), study week number (week_nr), session rating of perceived exertion (srpe), total distance covered (gps_td), high-speed running distance (gps_v4), sprint speed distance (gps_v5) and player load (gps_pl).

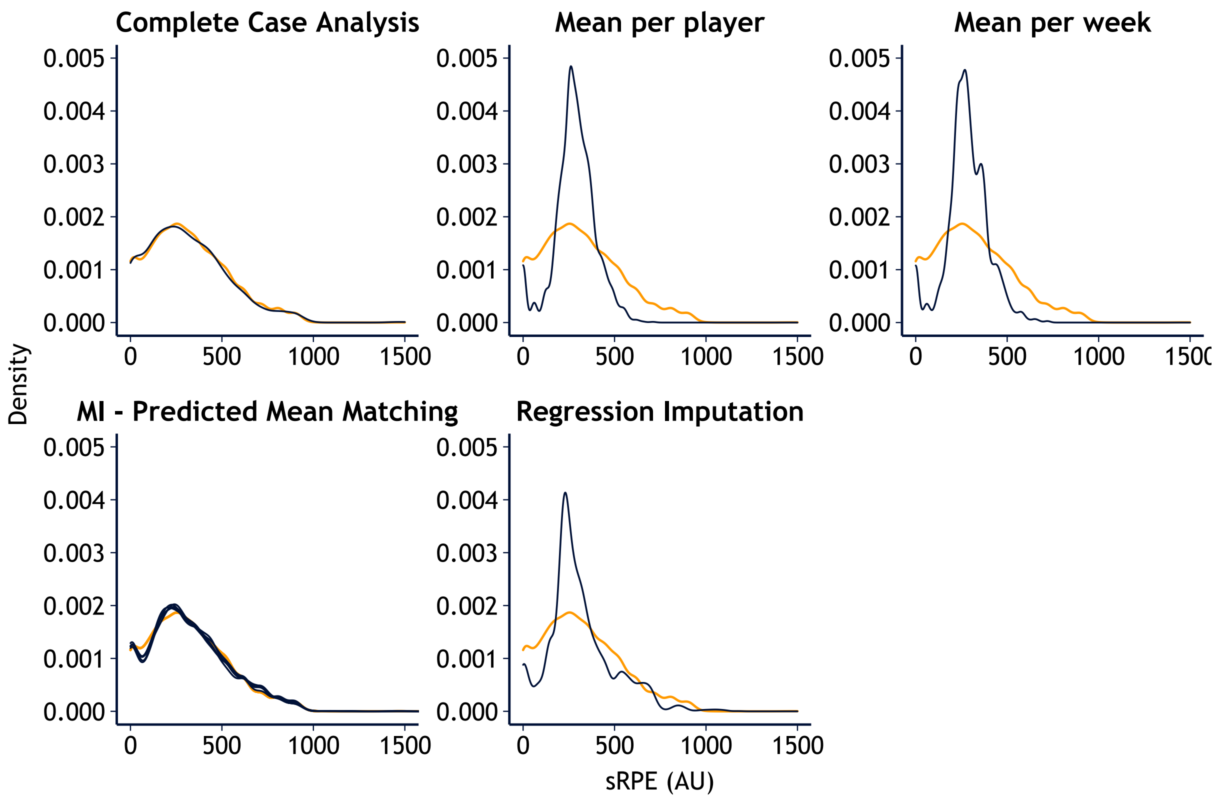

Figure S3. Distribution of session Rating of Perceived Exertion (sRPE) measured in Arbitrary Units (AU) after handling missing observations with five different methods. Blue lines are imputed values, yellow lines are the real data. The amount of missing data was set to 50% under the assumption of missing completely at random, meaning there was no systematic pattern in the missing data. Complete Case Analysis and Multiple Imputation (MI) using predicted mean matching were the only methods that managed to accurately retain the original distribution.

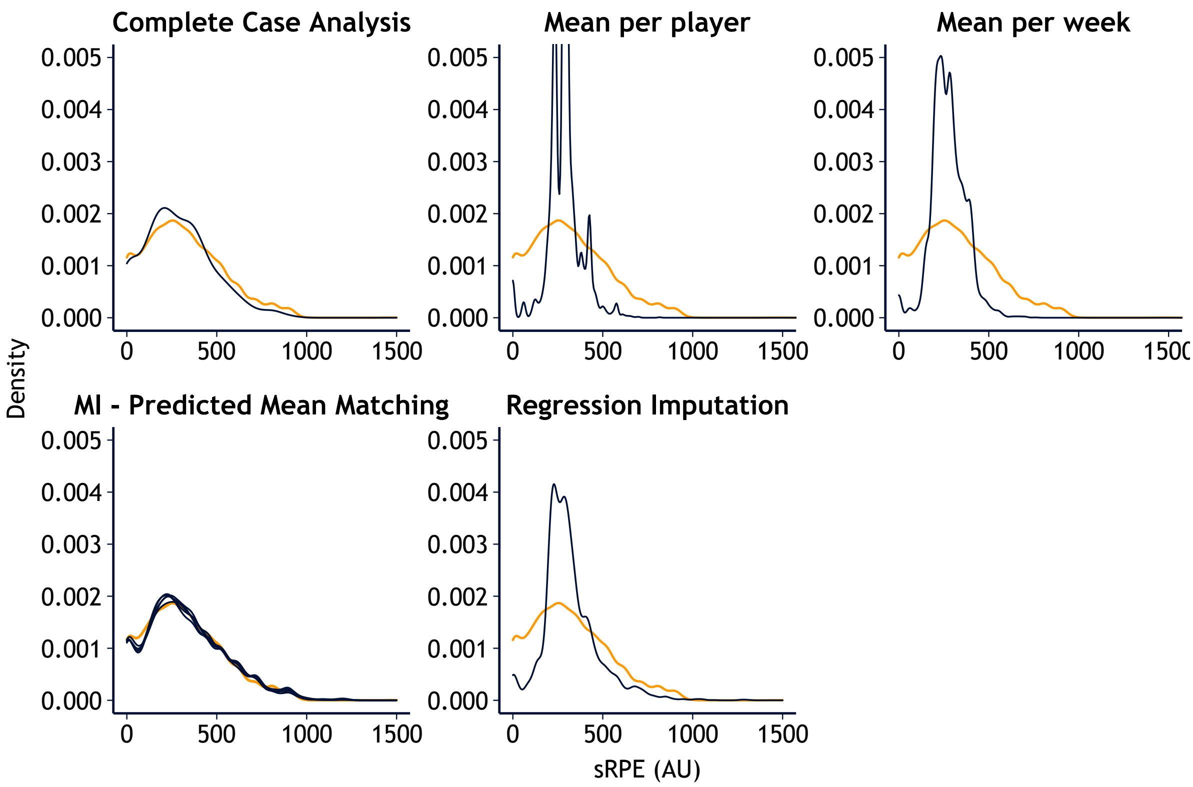

Figure S4. Distribution of session Rating of Perceived Exertion (sRPE) measured in Arbitrary Units (AU) after handling missing observations with five different methods. Blue lines are imputed values, yellow lines are the real data. Missing were introduced through a probability function under missing at random, meaning the probability of missing was dependent on other variables in the dataset. Around 80% missing was introduced under a strong relationship between the missing probability and the other variables. Complete Case Analysis and Multiple Imputation (MI) using predicted mean matching were the only methods that managed to accurately retain the original distribution.

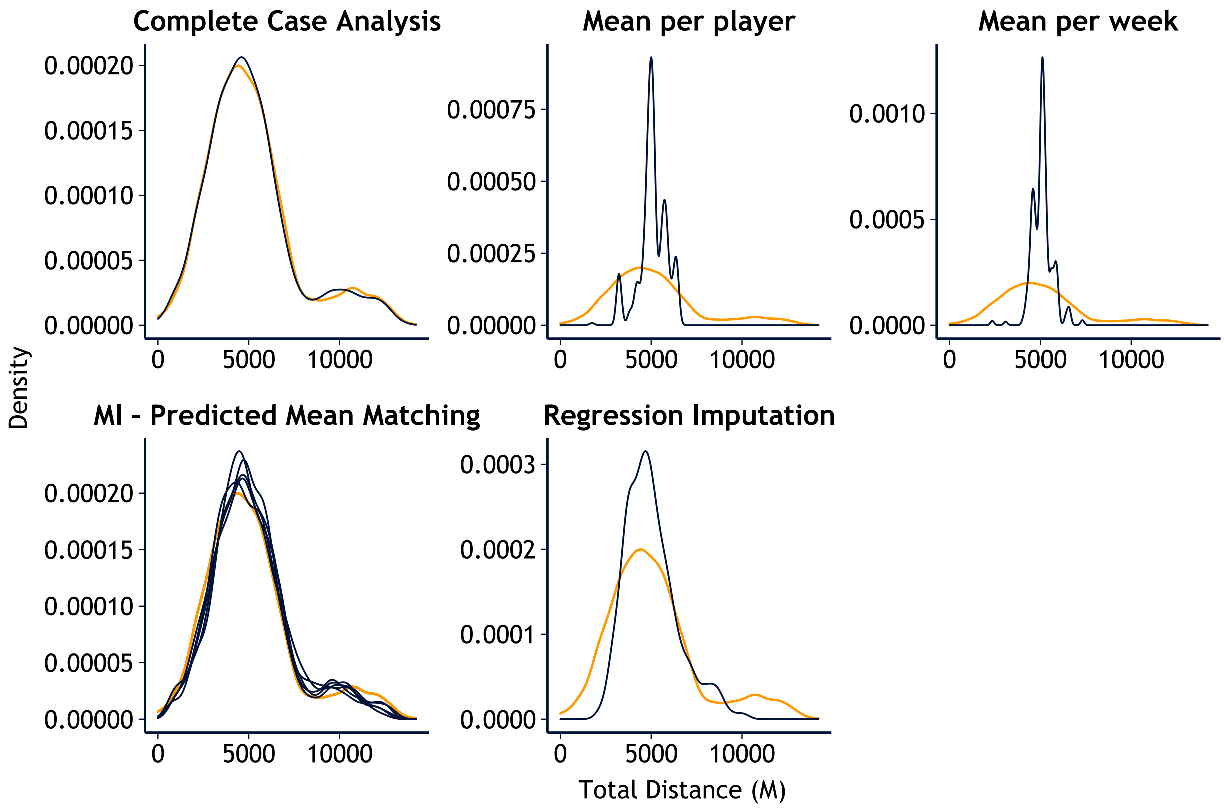

Figure S5. Distribution of total distance measured in meters (M) after handling missing total distance observations with five different methods. Blue lines are imputed values, yellow lines are the real data. The amount of missing data was set to 50% under the assumption of missing completely at random, meaning there was no systematic pattern in the missing data. In this case, if total distance was missing, all other GPS variables were also missing, and the player's position was among the variables in the imputation model. Complete Case Analysis and Multiple Imputation (MI) using predicted mean matching were the only methods that managed to accurately retain the original distribution.

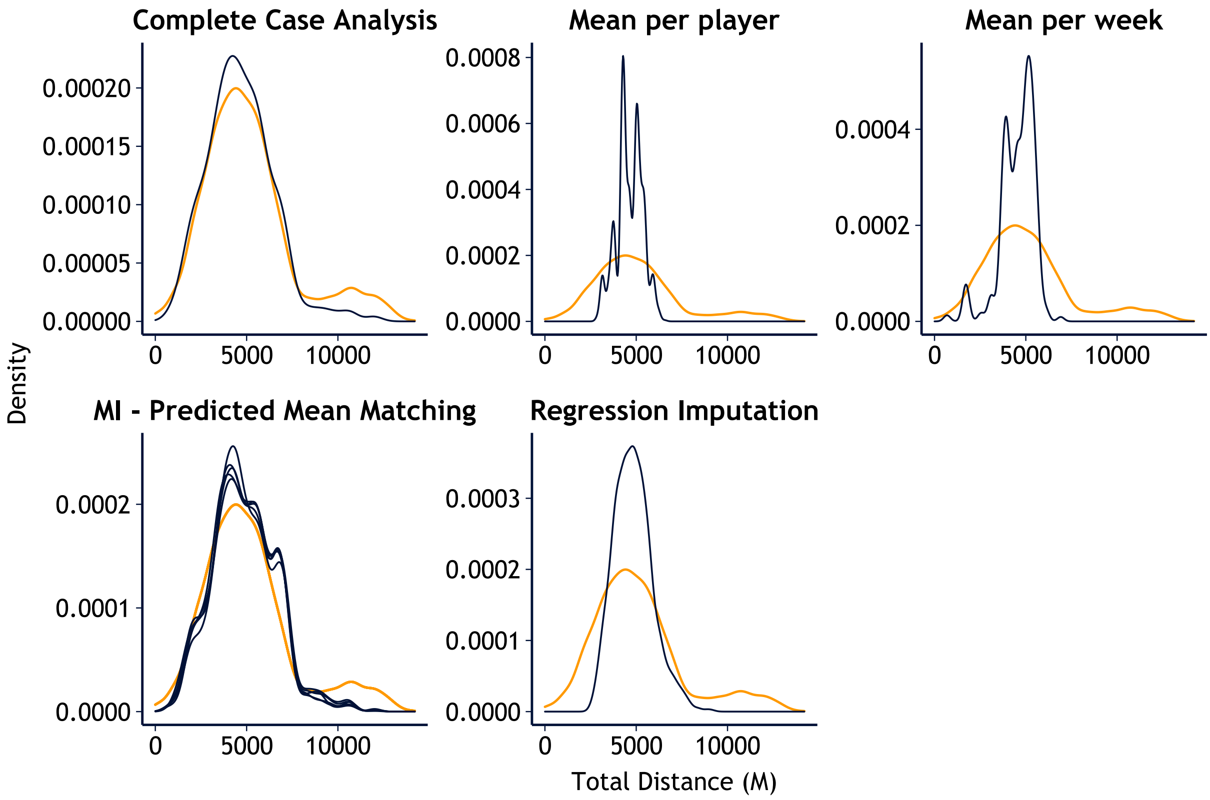

Figure S6. Distribution of total distance measured in meters (M) after handling missing total distance values with five different methods. Blue lines are imputed values, yellow lines are the real data. Missing were introduced through a probability function under missing at random, meaning the probability of missing was dependent on other variables in the dataset. Around 80% missing was introduced under a strong relationship between the missing probability and the other variables. In this case, if total distance was missing, all other GPS variables were also missing, and the player's position was among the variables in the imputation model. Complete Case Analysis and Multiple Imputation (MI) using predicted mean matching were the only methods that came close to retaining the original distribution.

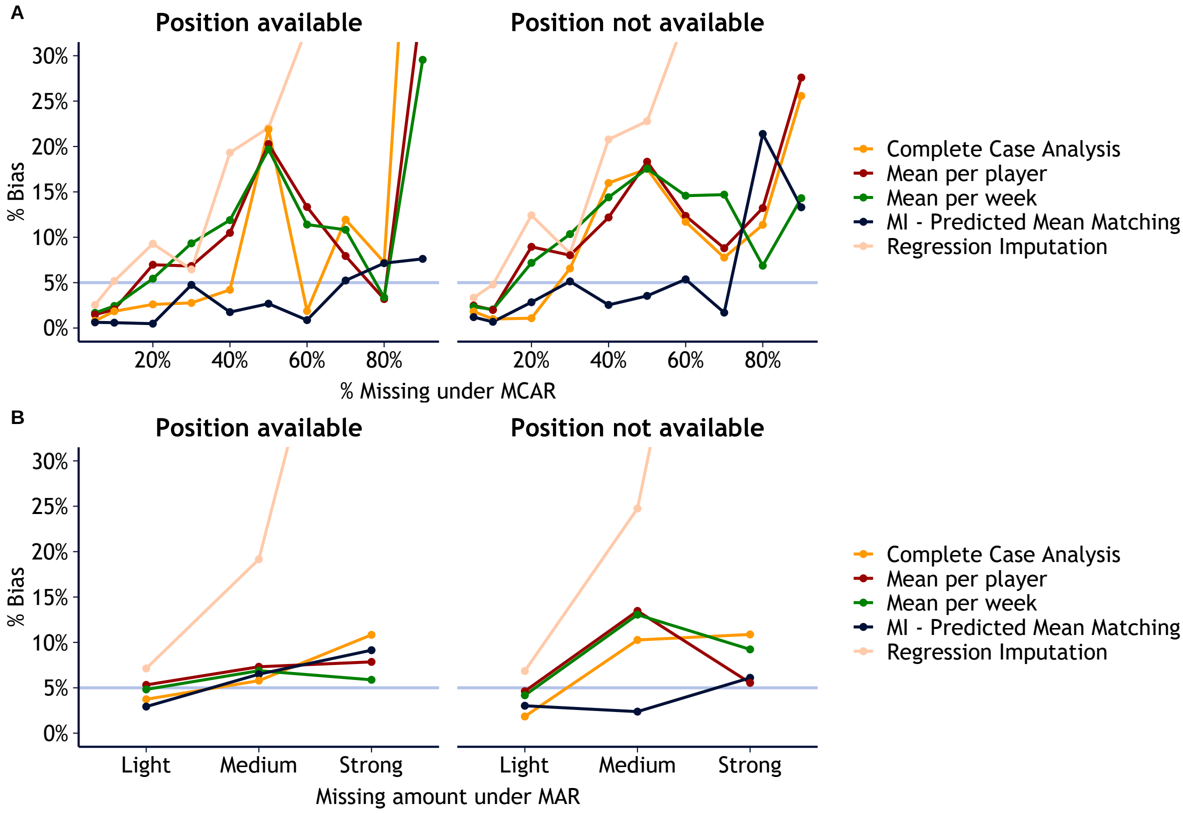

Figure S7. The performance of different methods of handling missing data in session Rating of Perceived Exertion when player position is among the variables in the dataset, and when it is not. Shown for (A) varying levels of missing data under the assumption of Missing Completely at Random (MCAR) and (B) varying levels of missing data under the assumption of Missing at Random (MAR). Regression imputation reaches off the chart to between 34% and 136% bias under MCAR and 85% under MAR.

Paper II

BMJ Open Sport & Exercise Medicine

Not straightforward: modelling non-linearity in training load and injury research

Lena Kristin Bache-Mathiesen [1], Thor Einar Andersen,[1]
Torstein Dalen-Lorentsen [1], Benjamin Clarsen,[1,2] Morten Wang Fagerland[1,3]

To cite: 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**:e001119. doi:10.1136/bmjsem-2021-001119

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjsem-2021-001119).

Accepted 16 July 2021

[1]Department of Sports Medicine, Oslo Sports Trauma Research Centre, Norwegian School of Sports Sciences, Oslo, Norway
[2]Centre for Disease Burden, Norwegian Institute of Public Health, Bergen, Norway
[3]Research Support Services, Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

Correspondence to
Lena Kristin Bache-Mathiesen; l.k.bache-mathiesen@nih.no

ABSTRACT

Objectives To determine whether the relationship between training load and injury risk is non-linear and investigate ways of handling non-linearity.

Methods We analysed daily training load and injury data from three cohorts: Norwegian elite U-19 football (n=81, 55% male, mean age 17 years (SD 1)), Norwegian Premier League football (n=36, 100% male, mean age 26 years (SD 4)) and elite youth handball (n=205, 36% male, mean age 17 years (SD 1)). The relationship between session rating of perceived exertion (sRPE) and probability of injury was estimated with restricted cubic splines in mixed-effects logistic regression models. Simulations were carried out to compare the ability of seven methods to model non-linear relationships, using visualisations, root-mean-squared error and coverage of prediction intervals as performance metrics.

Results No relationships were identified in the football cohorts; however, a J-shaped relationship was found between sRPE and the probability of injury on the same day for elite youth handball players ($p<0.001$). In the simulations, the only methods capable of non-linear modelling relationships were the quadratic model, fractional polynomials and restricted cubic splines.

Conclusion The relationship between training load and injury risk should be assumed to be non-linear. Future research should apply appropriate methods to account for non-linearity, such as fractional polynomials or restricted cubic splines. We propose a guide for which method(s) to use in a range of different situations.

Key messages

What is already known?

- ► Hypotheses suggest that the relationship between training load and injury risk is non-linear.
- ► Methods used in previous training load and injury research often assume linearity.
- ► Categorisation has been proven a suboptimal alternative for handling non-linearity.

What are the new findings?

- ► A non-linear relationship ($p<0.001$) between session rating of perceived exertion and the probability of injury in elite youth handball players would not have been discovered if linearity had been assumed ($p=0.24$).
- ► Acceptable Brier scores and C-statistics from a linear model do not mean that the relationship is linear.
- ► Categorising training load by quartiles could not model a linear relationship under skewed data conditions.
- ► Fractional polynomials and restricted cubic splines were the only methods capable of exploring non-linear shapes.

How might it impact clinical practice?

- ► Clinical researchers will have the tools available to perform causal and predictive research on training load and injury risk more accurately.
- ► More consistent methodology between training load and injury risk studies will improve comparability, reproducibility and facilitate meta-analyses.

INTRODUCTION

Injuries can hamper athlete and team performance in a variety of sporting disciplines.[1] Overuse injuries, in particular, are considered preventable, and in the last decade, researchers have investigated how training load affects injury risk in different football codes and other sports.[2] Results have been conflicting; some studies have found an increased risk with increased training loads, some have found that lower loads increase injury risk and some have found no association at all.[3 4] Hence, the relationship between training load and injury remains uncertain.

In 2013, Gamble theorised a U-shaped relationship between training load and injury risk. Too little and too much load increases risk,[5] with the middle section of the spectrum representing the lowest risk point. This hypothesis was revisited in 2016 by Blanch and Gabbett[6] who, based on three training load-injury datasets in different sports, postulated a workload–injury relationship that closely resembled a J-shaped curve; however, the statistical methodology in that paper has been questioned.[7] Gabbett[8] theorised a non-linear relationship between training load and injury risk with the rationale that training load may increase the risk of injury and build beneficial physiological adaptations such as aerobic capacity and strength, factors associated with decreased injury risk. The hypotheses of both Gamble and Gabbett suggest a non-linear relationship between different measures of training load and injury risk, prompting recent calls for better handling of non-linearity in the field.[9 10]

Despite these hypotheses and calls, methods that assume a linear relationship between training load and injury risk, such as Pearson correlations and logistic regression, are commonly used in the field.[11] If the training load and injury relationship is non-linear, such methods are expected to produce conflicting, irreproducible—and sometimes simply wrong—results. Nevertheless, no study has so far determined alternative methods for handling non-linearity.

The ideal method to handle non-linearity should be able to: (1) explore non-linear shapes and thus may confirm or reject previously outlined hypotheses; (2) model the non-linear relationship accurately; and (3) offer interpretable results.

The overall aim of this study was to identify the best methods for handling non-linearity in training load and injury research. First, we ascertained the relationship in three sports populations to reveal any potential evidence of non-linearity, to illustrate the problems and to present solutions. Second, we compared different methods in their ability to explore and accurately model potential non-linear shapes. Finally, we used the comparisons to develop a guide for which method(s) to use in different situations.

MATERIALS AND METHODS

Participants

We obtained training load and injury data collected from three cohorts: Norwegian elite U-19 football players (n=81, 55% male, mean age: 17 years, SD: 1 year),[12] one male football team from the Norwegian Premier League (n=36, mean age: 26 years (SD: 4))[13] and elite youth handball players recruited from Norwegian sports high schools (n=205, 36% male, mean age: 17 years (SD: 1)).[14] These cohorts were followed for 104, 323 and 237 days, respectively, during the competitive season.

All participants provided informed consent. Ethical principles were followed in accordance with the Declaration of Helsinki.

Training load definition

In all three cohorts, players reported the number of training sessions and matches daily. They also reported the duration of each activity and their rating of perceived exertion (RPE)[15] on the modified Borg CR10 scale.[16] To derive the session RPE (sRPE),[16] we multiplied the RPE by the activity duration in minutes.

Missing sRPE values are reported in online supplemental table S1 and were 24% for elite U-19 football, 41% for Premier League football and 64% for elite youth handball. The missing values were imputed using multiple imputation (online supplemental figure S1), a method that also performs well in cases of high amounts of missing (80%),[17] and the imputed values were deemed valid (online supplemental figure S2).

All load measures were based on players' daily ratings of perceived exertion (sRPE). We calculated an acute:chronic workload ratio (ACWR) in two different ways:

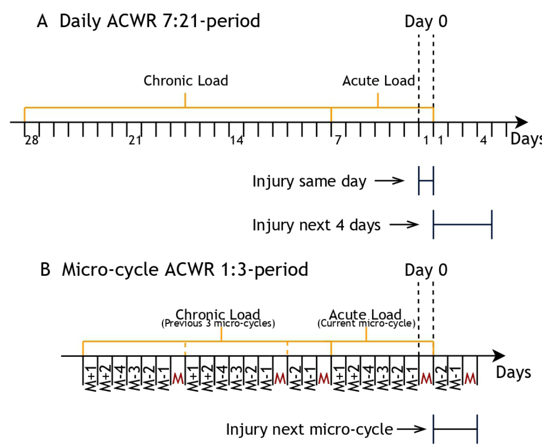

Figure 1 Illustration of time periods for calculating (A) daily ACWR 7:21-period and (B) micro-cycle ACWR 1:3-period. The first day that ACWR is calculated from is denoted day 0. The space between two tick marks represent 1 day (24 hours). For B, a microcycle period consists of all activity before a new match (M). That is, recovery days after the previous match as well as the training days before the next match. Days denoted with negative numbers are training days before the next match (M-1: being the day before the match; M-2: 2 days before a match and so on). Days with positive numbers are recovery and training days after a match (M+1: being the day after a match, M+2: 2 days after a match). The number of days between matches varies by the match schedule. How a team plan their training and recovery activities varies and is dependent on the teams' philosophy. For A, injury on the same day is defined as an injury on day 0, and future injury is defined as an injury occurring during the next 4 days excluding day 0. For B, future injury was defined as an injury occurring during the next microcycle excluding day 0. ACWR, acute:chronic workload ratio.

Daily ACWR 7:21-period

The mean sRPE across 7 days divided by the exponentially weighted moving average (EWMA) of the previous 21 days, uncoupled (figure 1).[18] The calculation was performed on a sliding window moving 1 day at a time from and including the 28th day.[19] The last day in the acute load is considered day 0 (figure 1).

Microcycle ACWR 1:3-period

The mean sRPE for each microcycle divided by the EWMA of the previous three microcycles uncoupled (figure 1). A microcycle was defined as all recovery days after the previous match and the training days before the next match. The next microcycle started on the first training day after the match and so on. For an illustration of a microcycle, see figure 1. The ACWR calculation was performed in the same manner as described for daily ACWR, on a sliding window moving one microcycle at a time from and including the fourth microcycle. The last day of the fourth microcycle was considered day 0 (figure 1).

When computing a ratio, one assumes that there is no relationship between the ratio and the denominator after controlling for the denominator; a ratio is only effective when the relationship between the numerator and the denominator is a straight line that intersects the origin.[20] Tests of this assumption are reported in online supplemental figure S3.

Injury definition

The same online questionnaire was used to collect daily health status and training information from all three sports cohorts. The elite U-19 football data and elite youth handball data were collected via the Briteback AB online survey platform, while the Norwegian Premier League football data were collected with Athlete Monitoring, Moncton. The players daily reported whether they had experienced 'no health problem', 'a new health problem' or 'an exacerbation of an existing health problem'. In the youth elite handball study, if players reported any new health problems, they were immediately prompted to specify whether it was an injury or illness in the questionnaire. In the football studies, if players reported any new health problems, a clinician contacted them by telephone the following day for a structured interview and classified the health problem as an injury or illness with the UEFA guidelines.[21] Players were asked to report all physical complaints, irrespective of their consequences on sports participation or the need to seek medical attention.[22]

Statistical analyses

To estimate the relationship between training load and injury risk, mixed effects logistic regression was used.[11 23]

We considered two outcomes: (1) occurrence of an injury on the same day as the observed training load (day 0) and (2) occurrence of injury in the future, where the current observation day (day 0) was not included. For unmodified training load values and daily ACWR 7:21-period, the future injury was defined as an injury occurring during the next 4 days excluding day 0. For microcycle ACWR 1:3-period, the future injury was any injury occurring during the next microcycle excluding day 0 (see figure 1 for an illustration of injury time periods and online supplemental table S2 for a list of the different models).

We adjusted for player age in all analyses. In addition, we adjusted for sex in the U-19 elite football and the elite youth handball models. In all models, the relationship between sRPE and injury risk was modelled with restricted cubic splines (RCSs).[24] The models were repeated without splines to simulate the relationship we would have discovered if we had assumed linearity. When using RCS, the estimated regression coefficients do not have a clinically meaningful interpretation, and only their p values are numerically interpretable.[24] The main result is, therefore, a visualisation of the model predictions (with uncertainty) to determine the shape of the relationship between training load and injury risk.

More details about data preparation and calculations are available in a supplementary file in .pdf format (online supplemental file 2). Our analyses served to illustrate whether there is any evidence for non-linearity in training load and injury research and should not be interpreted as causal inference.

Simulations

In addition to analysing real data, we performed (stochastic) simulations to compare different methods for ascertaining non-linear and linear relationships between training load and injury risk. The simulations were based on the elite U-19 football dataset since it had the least missing data (24%). The methodology here is focused on a causal research setting; however, the methods may also be applied in predictive research.[25] A detailed description of the simulation process and equations, as well as justifications for our methodological choices, is available as supplementary material (online supplemental file 2).

Two datasets were created. The first kept the original 8495 sRPE and 6308 ACWR values. In the second, sRPE and ACWR were sampled with replacement to generate 22 500 training load values.

Artificial injuries were simulated under different assumed scenarios for the relationship between training load and injury risk:

1. A U shape.
2. A J shape.
3. A linear shape.

A U shape between training load and injury risk indicates that the injury risk at lower levels of training load is equal to the injury risk at higher levels of training load. In contrast, moderate levels of training load have the lowest risk. In a J shape, moderate levels of training load have the lowest injury risk, followed by low levels of training load having intermediate risk. Finally, high levels of training load have the highest injury risk. For the U and linear relationship shapes, the simulated probability of an injury was based on the sRPE, while for the J shape, it was based on the ACWR. Any reference to the 'true' probability refers to the simulated probability we have created for a given scenario and which we aim to model.

We used mixed effects logistic regression models to estimate the relationship between training load and predefined injury risk, and we compared seven different methods to model the relationship:

- Linear model.
- Categorising by quartiles (data driven).
- Categorising by subjective cut-offs (subjective).
- Quadratic model.

- Fractional polynomials.
- RCSs with automated knots (data driven).
- RCSs with subjectively placed knots (subjective).

The root-mean-squared error (RMSE), coverage of prediction intervals, Brier score for model fit and C-statistics for predictive ability were calculated as performance measures. RMSE is a combined measure of accuracy and precision, where the lower the RMSE, the better the method. RMSE is only interpretable by comparing values in the same analysis – the values are meaningless in isolation.[26]

In summary, the four steps of the simulations were:

1. Sample training load values from the elite U-19 football data.
2. Simulate injuries with three different shapes for the relationship between injury risk and training load.
3. Fit seven different models with injury as the outcome and training load as the explanatory variable.
4. Calculate performance measures.

Steps 1–4 were repeated 1900 times.

For the U-shaped relationship, predicted values were visualised alongside the predefined shape to determine each method's ability to capture the true relationship. RMSE was also visually compared for the non-linear shapes.

All statistical analyses and simulations were performed using R V.4.0.2.[27] A GitHub repository is available with R code and data files.[28]

RESULTS

Evidence of non-linearity in training load and injury risk relationship research

A strong J-shaped relationship was found between sRPE and the probability of injury on the same day for elite youth handball players ($p<0.001$, figure 2A, online supplemental table S3). The linear model did not find this relationship (OR=1.0, 95% CI 0.99 to 1.00, p=0.24, figure 2B, online supplemental table S4). Additionally, for the handball cohort, an uncertain ∩-shaped relationship was present between sRPE and probability of injury in the next 4 days (p=0.06, figure 2B). These results also conflicted with the linear model showing no relationship (OR=1.0, 95% CI 0.99 to 1.00, p=0.35, figure 2B). For microcycle ACWR, the assumption that the relationship between the numerator and the denominator is a straight line intersecting the origin was supported, while for daily ACWR, the assumption was violated (online supplemental figure S3). No other relationships had significant p values or practically notable effect sizes (online supplemental table S3, figure S5 and S6).

Simulations

The quadratic model, fractional polynomials (FPs) and RCSs with subjectively placed knots were the only methods capable of modelling the non-linear U-shaped relationship (figure 3). FPs and RCS with subjectively placed knots (RCS subjectively) had the lowest RMSE and were, therefore, the best methods for the U shape (figure 4A).

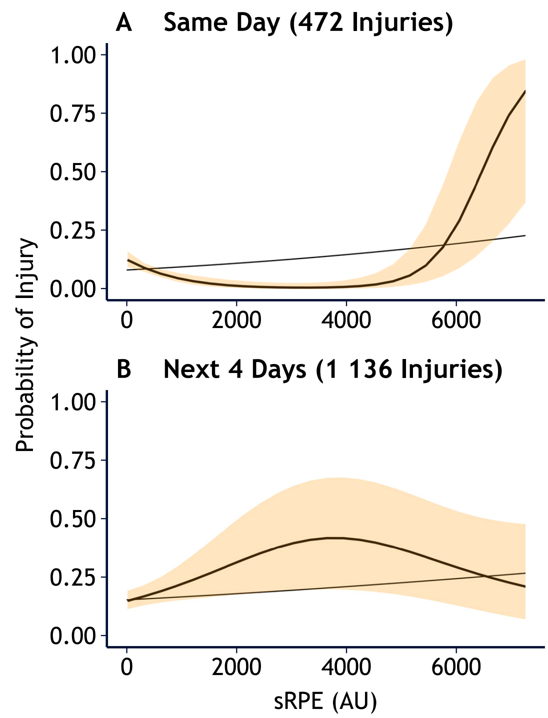

Figure 2 Probability of injury in elite youth handball on (A) the sameday and (B) the next 4 days, for each level of session rating of perceived exertion (sRPE) measured in arbitrary units (AU), as predicted by mixed effects logistic regression models with restricted cubic splines. The predictions pertain to a 17-year-old female. The yellow area represents 95% cluster-robust CIs around predicted values. The straight line shows the same predictions from an equivalent model without splines (ie, assuming linearity). For figure part B, modelling the response of injury in the next 4 days, multiple injuries on the same day were considered one event and an injury event would pertain to four load values and are therefore included four times.

The linear model had—by far—the highest RMSE and the data-driven RCS the second highest (figure 4A). In contrast, RCS (subjectively) had among the highest RMSE (figure 4B) regarding the J-shaped relationship. For the J shape, FPs and the quadratic model were the best methods (figure 4B). FPs had second-to-lowest RMSE for non-linear relationships (figure 4) and consistently had the best coverage (table 1).

All methods had a similar degree of error, predictive ability and model fit for the linear relationship (table 1).

The categorisation methods had the lowest coverage for the U and linear shapes, and categorising by quartiles had particularly poor coverage for the linear shape (25% vs >99% for other methods, table 1). For the J shape, the linear model performed worse than categorisation with 55% (vs 79% and 89%) for n=6308 (table 1). Predictions from the linear model could not form the U

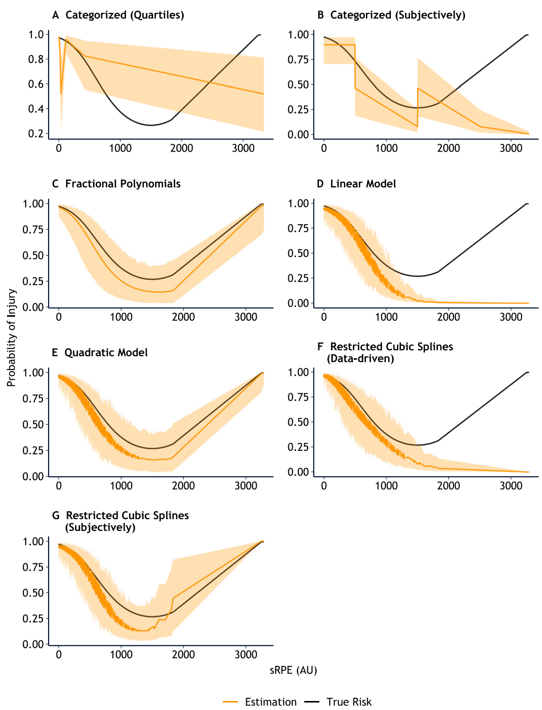

Figure 3 Probability of injury for each level of session rating of perceived exertion (sRPE) as predicted by seven different methods of modelling load. The yellow line represents the ability of the method to capture the U-shaped relationship (shown by the black line). The yellow area corresponds to the prediction interval. The predictions are based on 8494 sRPE values sampled from a highly skewed distribution in a Norwegian elite U-19 football dataset.

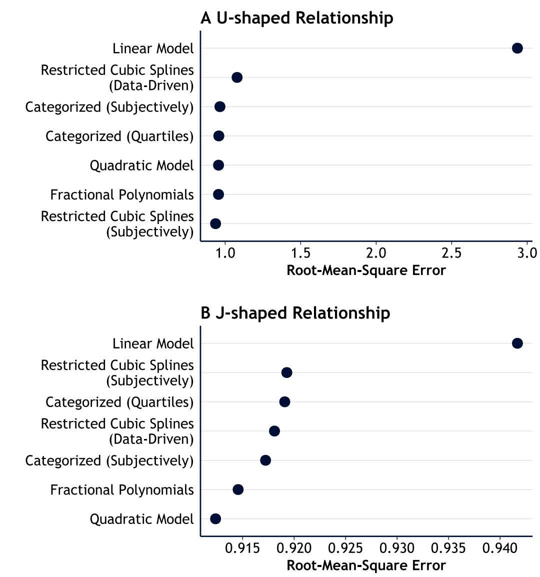

Figure 4 The mean root-mean-squared error (RMSE) of 1900 permutations for seven different methods modelling a non-linear (A) U-shaped relationship between session rating of perceived exertion (sRPE) and probability of injury, and (B) J-shaped relationship between acute:chronic workload ratio (ACWR) and probability of injury. The methods are arranged from top-to-bottom by the method with highest RMSE (most error) to the method with lowest RMSE. Thus, the best methods (those with lowest RMSE) are arranged towards the bottom. For figure part A, fractional polynomials and restricted cubic splines (subjectively) were the best methods, while for figure part B, fractional polynomials and the quadratic model were the best methods. The calculations are based on a Norwegian elite U-19 football dataset with 8494 sRPE values for (A) U shape and 6308 ACWR values for (B) J shape. RMSE cannot be compared between the two shapes, only within each shape.[26]

shape (figure 3) and had the highest degree of error for both non-linear shapes (highest RMSE; table 1, figure 4) but showed high predictive ability for the U shape (C-statistic >0.8) and moderate to poor predictive ability of the J shape (C-statistic=0.77 for n=6308, C-statistic=0.62 for n=22 500) in line with the other methods (table 1).

The differences in evaluation metrics between the two different sample sizes, n=22 500 and n=8494 for sRPE, and n=22 500 and n=6308 for ACWR, were negligible (table 1). Model fit determined by Brier score also failed to notably differentiate methods (table 1).

DISCUSSION

This is the first study exploring the potential for non-linearity in the relationship between training load and injury risk for football and handball. We found a J-shaped relationship between training load measured as the sRPE and probability of an injury on the same day in an elite youth handball cohort (figure 2A).

We also found that three methods were able to model the non-linear relationships between training load and injury explored in this paper: the quadratic model, FPs and RCSs, which managed to accurately recreate all simulated risk shapes (figure 4).

Evidence of non-linearity in training load and injury risk relationship research

All modelled relationships between training load and injury risk were either flat (no relationship) or non-linear. The results showed that the strength and direction of the relationship varied between training load—and injury—definitions in the handball population, while no relationships were found in the two football populations.

If we had assumed linearity and modelled the data accordingly, we would not have discovered these relationships. More grievously, we would have concluded there was no relationship between training load and injury risk for elite youth handball players for injury on the same day (linear model, p=0.24, type II error), when

Table 1 A comparison of mean root-mean-squared error, Brier score, C-statistic and coverage of prediction intervals for 1900 permutations of modelling the relationship between training load and risk of injury in seven different ways, with predetermined relationship shapes

Relationship	Sample size	Method	RMSE	Brier score	C-statistic	Coverage (%)
U shape	22 500	Linear model	2.344	0.097	0.827	100.000
		Categorised (quartiles)	0.995	0.101	0.809	99.678
		Categorised (subjectively)	0.996	0.102	0.758	94.600
		Quadratic model	0.993	0.097	0.826	100.000
		Fractional polynomials	0.994	0.096	0.829	100.000
		Restricted cubic splines (data driven)	1.065	0.097	0.826	100.000
		Restricted cubic splines (subjectively)	0.981	0.097	0.827	100.000
	8494	Linear model	2.935	0.093	0.851	98.048
		Categorised (quartiles)	0.958	0.096	0.838	98.769
		Categorised (subjectively)	0.965	0.098	0.809	84.600
		Quadratic model	0.956	0.092	0.850	98.937
		Fractional polynomials	0.956	0.092	0.852	98.942
		Restricted cubic splines (data driven)	1.079	0.092	0.849	98.686
		Restricted cubic splines (subjectively)	0.936	0.092	0.851	98.687
J shape	22 500	Linear model	1.044	0.063	0.618	77.694
		Categorised (quartiles)	0.993	0.064	0.689	88.652
		Categorised (subjectively)	0.993	0.063	0.690	96.404
		Quadratic model	0.984	0.061	0.732	99.997
		Fractional polynomials	0.986	0.061	0.740	100.000
		Restricted cubic splines (data driven)	0.992	0.061	0.735	99.999
		Restricted cubic splines (subjectively)	0.993	0.061	0.721	99.869
	6308	Linear model	0.942	0.060	0.774	54.493
		Categorised (quartiles)	0.919	0.060	0.791	79.120
		Categorised (subjectively)	0.917	0.059	0.795	89.393
		Quadratic model	0.912	0.057	0.817	93.272
		Fractional polynomials	0.915	0.057	0.821	95.517
		Restricted cubic splines (data driven)	0.918	0.057	0.818	94.281
		Restricted cubic splines (subjectively)	0.919	0.057	0.812	89.959
Linear	22 500	Linear model	0.999	0.239	0.591	100.000
		Categorised (quartiles)	0.999	0.240	0.588	25.000
		Categorised (subjectively)	0.999	0.241	0.579	99.995
		Quadratic model	0.999	0.239	0.591	99.999
		Fractional polynomials	0.999	0.239	0.592	100.000
		Restricted cubic splines (data driven)	0.999	0.239	0.591	100.000
		Restricted cubic splines (subjectively)	0.999	0.239	0.591	99.997
	8494	Linear model	0.991	0.228	0.655	99.795
		Categorised (quartiles)	0.991	0.228	0.653	24.957
		Categorised (subjectively)	0.991	0.229	0.649	99.678
		Quadratic model	0.991	0.228	0.656	99.786
		Fractional polynomials	0.991	0.228	0.656	99.788
		Restricted cubic splines (data driven)	0.991	0.228	0.656	99.789
		Restricted cubic splines (subjectively)	0.991	0.228	0.656	99.791

RMSE, root-mean-squared error.

it was, in fact, a strong U-shaped parabola (RCS model, p<0.001, figure 2A). This may happen when a relationship is not only non-linear but non-monotonic. In monotonic relationships, the response variable Y (injury probability) moves only in one direction as X (training load) increases, while in non-monotonic relationships, Y sometimes increases and sometimes decreases when X increases.[9]

In 2013, Gamble[5] theorised a U-shaped relationship between training load and risk of injury. Data presented by Blanch and Gabbett[6] suggested a J-shaped relationship between ACWR and injury, although the methodology and interpretation of this finding have recently been questioned.[7] Here, we reproduced a J shape between sRPE and injury occurring on the same day for elite youth handballers but not for the relative training load described by the ACWR in the same cohort. In Lathlean *et al*,[29] a U shape was discovered between training load and the risk of future injury in an Australian football cohort. These findings might suggest that the training load and injury relationship is different for different sports and populations. Since non-linearity is possible in a training load and injury context, we recommend assuming the data have an unknown, non-linear relationship when conducting statistical analyses.

Methods for addressing non-linear relationships

As expected, standard logistic regression could not model the U and J shapes, as it assumes linearity. For the U shape, the RMSE was threefold higher for the linear model than all other models (RMSE=2.9 vs RMSE≈0.95, figure 4A), showing that violation of the linearity assumption causes major bias and can substantially alter conclusions based on the results. Misleadingly, the linear model had a great C-statistic score (>0.8) and comparable Brier scores. This happened because the sRPE values were highly skewed (online supplemental figure S4). Over 90% of the data points were congested in the left-hand side of the U shape (figure 3, online supplemental figure S4). The linear model, which only managed to model the left-hand side of the U shape, therefore predicted most of the values well, causing the impressive C-statistic. However, it could not predict the right-hand side of the U shape at all and therefore had high RMSE. Consequently, a researcher who measures model fit by predictive ability alone may be falsely assured that the linearity assumption holds true.

Categorisation has previously been explored thoroughly in Carey *et al*[30] and proven a poor method for modelling non-linear relationships. The results were reproduced in our study using a football population, where the RMSE and coverage for categorisation were consistently outperformed by other methods (table 1). In addition, our results showed that categorising by quartiles was suboptimal for modelling non-linear relationships and also suboptimal when the relationship between training load and injury risk was linear (coverage of 25% vs >99% for all other methods).

Recently, some studies have added a quadratic term to the training load and injury model to test for linearity: if the term was non-significant, it was discarded for a linear model; if significant, they categorised the training load variable to handle non-linearity.[31–33] If the quadratic term is significant, the researchers correctly choose other options over a linear model. However, the quadratic term only tests for a parabolic shape—not non-linearity in general. A significant quadratic term does not mean the relationship is quadratic (parabolic). It means that a quadratic shape fits better than a linear shape. If the quadratic term is not significant, it does not necessarily mean the underlying relationship is linear, either, only that a quadratic shape fits poorly. Furthermore, testing non-linearity with a quadratic term has been shown to inflate type I error rates by 50%.[34]

Blanch and Gabbett[6] and Carey *et al*[19] used quadratic regression assuming a parabolic relationship between training load and injury risk. In our study, quadratic regression modelled the U-shaped risk profiles with low degrees of error (figures 3 and 4A) and had the best performance for the J-shaped relationship (figure 4B). This is expected, as the J shape was initially constructed from a quadratic model in Blanch and Gabbett.[6] Contrary to a real-life setting, however, we knew the risk profiles before analysing our data. Quadratic regression does not explore shapes but constrains the model to follow a specific pathway. We think it is only appropriate when strong evidence from previous studies support a parabolic relationship. We recommend assuming non-linearity of unknown shapes and using methods not to test for linearity but to explore and model non-linearity to discover the relationship. Based on our findings and previous research in other fields such as medical statistics,[35] FPs and RCSs appear to be the best methods for doing this.

FPs modelled all risk shapes accurately (figure 4, table 1). FP has recently been used in a training load and injury risk study.[29] This method requires minor subjective influence, and the results are intuitive, especially for users familiar with quadratic regression. Although it appears the superior choice at first glance, the method has a disadvantage: FPs are defined only for positive values, which means that an FP model is unable to model negative values and the value 0. In the context of training load and injury risk research, training load is (traditionally) never measured on a negative scale.[36] If it can be justified, adding a small constant (such as 0.001, or whatever is considered small in the context of the measuring scale) to all training load values can solve the problem with 0 and allow the use of FPs.

RCSs performance depended on how knot locations (the points where the polynomials that make up cubic splines are joined, see online supplemental file 2 for details) were chosen. In the data-driven method, where knots were automatically placed by the default setting, RCS failed to model the U-shaped scenario (figure 3). When knot position was chosen based on the range of

Box 1 Recommended methods to model non-linear relationships between training load and injury risk

To model non-linear relationships, either Fractional Polynomials (FP) or Restricted Cubic Splines (RCS) can be used.
Fractional polynomials are easier to interpret. We recommend FP under the following conditions:

- When the main objective is causal research, FP is preferred. When the training load measure does not include negative numbers or 0. This includes:
  - Studies that use the Acute-Chronic Workload Ratio or other metrics that cannot be the value 0 or a negative value.
  - Studies that model the relationship between training load and injury risk on the same day, or other scenarios where the researchers may wish to remove the days where the athletes were not exposed to any training load from the dataset.
  - Studies that can justify applying a small constant (such as 0.001, or whatever is considered small in the context of the measuring scale) to all training load values.

We recommend restricted cubic splines under the following conditions:

- When the main objective is predictive research, RCS is preferred.
- When the training load measures must have the value 0. This includes studies that wish to capture a change in the effect, regardless how small, going from no training load at all to any amount of training load.
- When training load is included in the study merely to adjust for it as a potential confounder and is not the main variable of interest.

We do not recommend changing the study aims or the chosen measure to use FP, nor do we recommend using FP under certain conditions and RCS for other conditions in the same study.
A step-by-step guide to performing FP and RCS in R can be accessed on the primary author's GitHub page.[39 40]

the training load variable, RCS modelled the U accurately (figure 3). However, the results were the opposite for the J-shaped relationship where the data-driven method was among those of lowest error, and the subjectively located knots had the highest amount of error (figure 4B). The default placement algorithm was by quartiles, and in the highly skewed distribution of the sRPE values used in the U-shaped relationship (online supplemental figure S4), it caused the knots to be placed tightly together (figure 3). Therefore, it could not model the shape, while the subjective version was created with the range of the values in mind. The ACWR values used in the J shape had a Gaussian distribution (online supplemental figure S4), and using quartiles was a feasible choice. This shows the importance of careful model calibration using clinical knowledge and knowledge of the data.

RCS produces effect sizes that are difficult to use in a practical setting, and results can only be interpreted in the form of p values and visualisation (such as in figure 2). RCS is less ideal than FP in causal research. Still, its disadvantages are not as relevant in predictive research where interpretability is of minor concern.[25] We propose a guide for when FP is recommended and when RCS is recommended (box 1).

Limitations

A limitation of this study was the sample size, the number of injuries and consequently statistical power. Neither of the two football cohorts satisfied the recommendation of >200 injuries to detect a small to moderate effect.[37] The elite youth handball data, despite having a sufficient number of injuries, had high amounts of missing sRPE values (64%), and this may have caused selection bias. We emphasise that the exploration of non-linearity in these data were for illustrative purposes and not to show causal inference.

We used statistical methods commonly used and recommended in the field to demonstrate how non-linear relationships can be ascertained with existing methods. We were consequently limited in the choice of methods. The ACWR model is under debate, and the pros and cons of the method have been explored extensively in recent publications.[12 18 38] The purpose of this paper was not to provide additional insight into that discussion but rather to demonstrate how a continuous training load variable should be modelled to account for non-linearity. For this reason, we opted to use ACWR, as it is currently the most used training load method in the field of training load and injury risk research.[4]

CONCLUSION

Exploratory analyses showed evidence of a non-linear relationship between training load and risk of injury in a sports population. Researchers should assume that the relationship between training load and injury risk is non-linear and use appropriate methods that explore relationships rather than constrain them. Linear methods should only be used when the relationship is first proven to be linear. We promote FPs or RCSs to model non-linear relationships, depending on the scenario.

Twitter Lena Kristin Bache-Mathiesen @lena_kbm, Thor Einar Andersen @DocThorAndersen, Torstein Dalen-Lorentsen @torsteindalen, Benjamin Clarsen @benclarsen and Morten Wang Fagerland @FagerlandWang

Acknowledgements We would like to thank Christian Thue Bjørndal for access to the elite youth handball data. We would also like to thank Garth Theron for high-quality programming of the Norwegian Premier League football database. This research would not be made possible without the collaboration of coaches and athletes, and we would like to thank the participants who contributed.

Contributors LKB-M designed the study and performed statistical analyses in collaboration with and under supervision from MWF and TEA. TDL constructed the novel idea of using microcycles instead of calendar weeks. All authors contributed with notable critical appraisal of the text and approved the final version.

Funding The Oslo Sports Trauma Research Centre provided all funding for performing this study.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval Study protocol for all three studies were approved by the Norwegian Centre for Research Data: Norwegian elite U-19 football (5487), Norwegian Premier League football (722773) and Norwegian elite youth handball (407930). They were also approved by the Ethical Review Board of the Norwegian School of Sport Sciences. The Norwegian elite U-19 football study was also approved by the South-Eastern Norway Regional Committee for Medical and Health Research Ethics (2017/1015).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available on reasonable request. Data used for simulations are available in a public, open access repository (https://github.com/lenakba/load-injury-non-linearity-study). The Norwegian elite U-19 football data, Norwegian Premier League football data and Norwegian elite youth handball data are available on reasonable request. These are anonymised based on requirements of the Norwegian Data Protection Agency. The removal of background variables for the anonymisation renders the data unusable for any reproducibility purposes; the data are only available for the sake of transparency.

ORCID iDs
Lena Kristin Bache-Mathiesen http://orcid.org/0000-0002-9466-7003
Torstein Dalen-Lorentsen http://orcid.org/0000-0003-4062-7601

REFERENCES

1 Hägglund M, Waldén M, Magnusson H, *et al*. Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA champions League injury study. *Br J Sports Med* 2013;47:738–42.
2 Griffin A, Kenny IC, Comyns TM. The association between the acute: chronic workload ratio and injury and its application in team sports: a systematic review. *Sports Medicine* 2019:1–20.
3 Eckard TG, Padua DA, Hearn DW, *et al*. The relationship between training load and injury in athletes: a systematic review. *Sports Medicine* 2018;48:1929–61.
4 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 Med* 2020;50:1–23.
5 Gamble P. Reducing injury in elite sport-is simply restricting workloads really the answer? *N Zeal J Sports Med* 2013;36.
6 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. *Br J Sports Med* 2016;50:471–5.
7 Impellizzeri F, Wookcock S, McCall A. *The acute-chronic workload ratio-injury figure and its 'sweet spot'are flawed*, 2019.
8 Gabbett TJ. The training-injury prevention paradox: should athletes be training smarter *and* harder? *Br J Sports Med* 2016;50:273–80.
9 Stern BD, Hegedus EJ, Lai Y-C. State dependence: does a prior injury predict a future injury? *Phys Ther Sport* 2021;49:8–14.
10 Kalkhoven JT, Watsford ML, Coutts AJ. Training load and injury: causal pathways and future directions. *Sports Med* 2021:1–14. doi:10.1007/s40279-020-01413-6
11 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:e022626.
12 Dalen-Lorentsen T, Andersen TE, Bjørneboe J. A cherry tree ripe for picking: the relationship between the acute: chronic workload ratio and health problems. *J Orthop Sports Phys Ther* 2021;0:1–26.
13 Theron GF. *The use of data mining for predicting injuries in professional football players*. University of Oslo, 2020.
14 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. *Front Sports Act Living* 2021;3:635103.
15 Borg G, Hassmén P, Lagerström M. Perceived exertion related to heart rate and blood lactate during arm and leg exercise. *Eur J Appl Physiol Occup Physiol* 1987;56:679–85.
16 Foster C, Florhaug JA, Franklin J, *et al*. A new approach to monitoring exercise training. *J Strength Cond Res* 2001;15:10Y 115.
17 Madley-Dowd P, Hughes R, Tilling K, *et al*. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol* 2019;110:63–73.
18 Wang C, Vargas JT, Stokes T, *et al*. Analyzing Activity and Injury: Lessons Learned from the Acute:Chronic Workload Ratio. *Sports Medicine*2020;50:1243–54.
19 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. *Br J Sports Med* 2017;51:1215–20.
20 Curran-Everett D. Explorations in statistics: the analysis of ratios and normalized data. *Adv Physiol Educ* 2013;37:213–9.
21 Hägglund M*et al*. Methods for epidemiological study of injuries to professional football players: developing the UEFA model. *Br J Sports Med* 2005;39:340–6.
22 Fuller CW, Ekstrand J, Junge A, *et al*. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Scand J Med Sci Sports* 2006;16:83–92.
23 Nielsen Rasmus Østergaard, Shrier I, Casals M, *et al*. Statement on methods in sport injury research from the first methods matter meeting, Copenhagen, 2019. *J Orthop Sports Phys Ther* 2020;50:226–33.
24 Durrleman S, Simon R. Flexible regression models with cubic splines. *Stat Med* 1989;8:551–61.
25 Shmueli G. To explain or to predict? *Statistical Science* 2010;25:289–310.
26 Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med* 2019;38:2074–102.
27 R Core Team. *R foundation for statistical computing, Vienna, Austria. 4.0.2*, 2013. http://www.R-project.org/
28 Bache-Mathiesen L. Modelling non-linearity in training load and injury research: R code Repository, 2021. Available: https://github.com/lenakba/load-injury-non-linearity-study [Accessed 12 Apr 2021].
29 Lathlean TJH, Gastin PB, Newstead SV, *et al*. Absolute and relative load and injury in elite junior Australian football players over 1 season. *Int J Sports Physiol Perform* 2019;1:1–9.
30 Carey DL, Crossley KM, Whiteley R, *et al*. Modeling training loads and injuries: the dangers of Discretization. *Med Sci Sports Exerc* 2018;50:2267–76.
31 Warren A, Williams S, McCaig S, *et al*. High acute:chronic workloads are associated with injury in England & Wales Cricket Board Development Programme fast bowlers. *J Sci Med Sport* 2018;21:40–5.
32 Sampson JA, Murray A, Williams S, *et al*. Injury risk-workload associations in NCAA American College football. *J Sci Med Sport* 2018;21:1215–20.
33 Ahmun R, McCaig S, Tallent J, *et al*. Association of daily workload, wellness, and injury and illness during Tours in international cricketers. *Int J Sports Physiol Perform* 2019;14:369–77.
34 Grambsch PM, O'Brien PC. The effects of transformations and preliminary tests for non-linearity in regression. *Stat Med* 1991;10:697–709.
35 Binder H, Sauerbrei W, Royston P. Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. *Stat Med* 2013;32:2262–77.
36 Bourdon PC, Cardinale M, Murray A, *et al*. Monitoring athlete training loads: consensus statement. *Int J Sports Physiol Perform* 2017;12:S2-161-S2-70
37 Bahr R, Holme I. Risk factors for sports injuries--a methodological approach. *Br J Sports Med* 2003;37:384–92.
38 Impellizzeri FM, Tenan MS, Kempton T, *et al*. Acute:Chronic Workload Ratio: Conceptual Issues and Fundamental Pitfalls. *Int J Sports Physiol Perform* 2020;15:907–13.
39 Bache-Mathiesen L. Performing Fractional Polynomials in R [PDF]. GitHub, 2021. Available: https://github.com/lenakba/load-injury-non-linearity-study/blob/main/guide_fractional_polynomials.pdf [Accessed 12 Apr 2021].
40 Bache-Mathiesen L. Performing Restricted Cubic Splines in R [PDF]. GitHub, 2021. Available: https://github.com/lenakba/load-injury-non-linearity-study/blob/main/guide_restricted_cubic_splines.pdf [Accessed 12 Apr 2021].

SUPPLEMENTARY RESULTS

Tables

Table S1. Data quality comparison of sports cohorts: the Norwegian elite U-19 football data (55% male, age; mean ± standard deviation (SD) = 17 ± 1 years), Norwegian Premier League football data (all male, age 26 ± 4 years) and elite youth handball data (36% male, age 17 ± 0.9 years).

		Football U-19	Football Elite	Handball
Sample Size	Number of athletes	81	36	205
	Number of sRPE values before imputation	6 424	6 061	17 268
	Number of sRPE values after imputation	8 495	10 232	47 651
	Number of injuries	81	38	472
	Number of injuries per athlete, mean (SD)	1 (1.2)	1 (1.5)	2.3 (2.9)
Missing data	Missing load values, n (%)	2 071 (24%)	4 171 (41%)	30 383 (64%)
	Missing load values per athlete, mean (SD)	26 (32)	116 (62)	148 (71)
Timelines	Mean (SD) answering time, days	0.3 (0.7)	0.01 (0.2)	0.7 (1.6)
	Percentage of forms answered the same day	72%	99%	53%
	Max answering time, days	9	4	119

Abbreviations: Football Elite, Norwegian Premier League; sRPE, session Rating of Perceived Exertion

Bache-Mathiesen LK, *et al. BMJ Open Sp Ex Med* 2021; 7:e001119. doi: 10.1136/bmjsem-2021-001119

Table S2. Overview of injury definition and models run on each sport population, with the number of load values and the number of injuries used in each model.

Population	Injury Definition[1]	Load Definition[2]	Load Values (n)[3]	Injuries (n)[3]
Football U-19 (n = 81)	Same day	sRPE	8495	81
		Daily ACWR 7:21-period	6308	43
	Next 4 days	sRPE	8495	210
		Daily ACWR 7:21-period	6308	129
	Next micro-cycle	Micro-cycle ACWR 1:3-period	793	26
Football Elite (n = 36)	Same day	sRPE	10 232	38
		Daily ACWR 7:21-period	9 260	32
	Next 4 days	sRPE	10 232	44
		Daily ACWR 7:21-period	9 260	34
	Next micro-cycle	Micro-cycle ACWR 1:3-period	553	26
Handball (n = 205)	Same day	sRPE	47 651	472
		Daily ACWR 7:21-period	42 116	320
	Next 4 days	sRPE	47 651	1 136
		Daily ACWR 7:21-period	42 116	714
	Next micro-cycle	Micro-cycle ACWR 1:3-period	1 897	242

Abbreviations: ACWR, Acute: Chronic Workload Ratio; Football Elite, Norwegian Premier League; sRPE = daily session Rating of Perceived Exertion; TL, Training Load.

[1]Same day was injury same day as the measured load value; Next 4 days was one or more injuries during the four days after the measured load value; Next micro-cycle was one or more injuries during the micro-cycle after the micro-cycle of the measured load values.

[2]Daily ACWR 7:21-period was the 7-day acute sRPE divided by previous 21-day chronic sRPE per day; Micro-cycle ACWR 1:3-period was the 1-micro-cycle acute sRPE divided by previous 3-micro-cycle chronic sRPE per micro-cycle. A micro-cycle was defined as all recovery days after the previous match as well as the training days before the next match.

[3]Due to aggregations, ACWR calculations and injury time-windows, the number of load values and injury events varied between models.

Table S3. Odds ratio with 95% confidence intervals, standard error, degrees of freedom and p-values from modelling the relationship between training load and injury risk using mixed effect models with restricted cubic splines.

Population	Load Definition[1]	Injury Definition[2]	Variable	OR[3]	CI 2.5%	CI 97.5%	SE	df	p
Football U-19	sRPE	Same day	Intercept	0.004	<0.001	0.939	2.808	941	0.047
			Load	1.000	0.997	1.003	0.002	3331	0.837
			Load'	1.001	0.997	1.004	0.002	3337	0.746
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.163	0.628	2.155	0.314	3286	0.631
			Age (Years)	1.088	0.800	1.479	0.157	921	0.592
		Next 4 days	Intercept	0.031	<0.001	13.614	3.094	273	0.262
			Load	1.001	1.000	1.002	0.001	4253	0.179
			Load'	0.999	0.998	1.001	0.001	3386	0.502
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.181	0.582	2.400	0.362	4727	0.645
			Age (Years)	0.973	0.689	1.373	0.175	261	0.874
	Daily ACWR 7:21-period	Same day	Intercept	0.002	<0.001	1.073	3.233	1088	0.053
			Load	0.778	0.313	1.936	0.465	1896	0.589
			Load'	2.970	0.586	15.057	0.827	1268	0.189
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.326	0.648	2.716	0.365	2592	0.440
			Age (Years)	1.118	0.785	1.594	0.181	1131	0.536
		Next 4 days	Intercept	<0.001	<0.001	25.567	6.179	104	0.148
			Load	4.285	1.241	14.793	0.631	498	0.021
			Load'	0.032	0.007	0.139	0.745	1565	<0.001
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.278	0.381	4.283	0.617	2328	0.691
			Age (Years)	1.160	0.594	2.266	0.338	110	0.661

Continues on the next page

Population	Load Definition[1]	Injury Definition[2]	Variable	OR[3]	CI 2.50 %	CI 97.50 %	SE	df	p
Football U-19	Micro-cycle ACWR 1:3-period	Next micro-cycle	Intercept	0.041	<0.001	40.086	3.502	396	0.362
			Load	0.210	0.012	3.536	1.439	562	0.278
			Load'	8.535	<0.001	6136037	6.857	356	0.755
			Load''	0.144	<0.001	3.52E+20	25.026	296	0.938
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.107	0.478	2.563	0.427	641	0.812
			Age (Years)	1.070	0.732	1.563	0.193	358	0.726
Football Elite	sRPE	Same day	Intercept	0.001	<0.001	0.011	1.437	4480	<0.001
			Load	1.000	0.995	1.005	0.003	4479	0.897
			Load'	1.001	0.994	1.008	0.004	4475	0.847
			Age (Years)	1.096	0.997	1.204	0.048	4480	0.056
		Next 4 days	Intercept	<0.001	<0.001	0.022	2.581	1593	0.001
			Load	0.998	0.994	1.003	0.002	168	0.501
			Load'	1.004	0.997	1.011	0.003	99	0.29
	Daily ACWR 7:21-period	Same day	Intercept	<0.001	<0.001	0.022	2.465	55	0.001
			Load	3.389	0.042	273.286	2.119	22	0.57
			Load'	0.337	0.004	31.613	2.201	24	0.626
			Age (Years)	1.104	0.994	1.226	0.053	3833	0.064
			Age (Years)	1.186	0.991	1.418	0.091	1662	0.062
		Next 4 days	Intercept	<0.001	<0.001	0.015	3.485	300	0.002
			Age (Years)	1.202	0.978	1.477	0.105	1349	0.081
			Load	6.731	0.116	390.17	1.948	20	0.339
			Load'	0.056	0.001	5.583	2.252	29	0.21
	Micro-cycle ACWR 1:3-period	Next micro-cycle	Intercept	<0.001	<0.001	0.136	2.841	62	0.009
			Age (Years)	1.113	1.016	1.219	0.046	476	0.021
			Load	7.523	0.030	1881.323	2.742	46	0.466
			Load'	0.340	0.005	22.344	2.113	112	0.610

Continues on the next page

Population	Load Definition[1]	Injury Definition[2]	Variable	OR[3]	CI 2.50 %	CI 97.50 %	SE	df	p
Handball	sRPE	Same day	Intercept	0.083	0.003	2.711	1.777	1632	0.162
			Load	0.999	0.998	0.999	<0.001	9445	<0.001
			Load'	1.002	1.001	1.003	0.001	2603	<0.001
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.112	0.780	1.586	0.181	11867	0.556
			Age (Years)	0.963	0.787	1.177	0.102	1740	0.711
		Next 4 days	Intercept	0.606	0.007	54.891	2.297	1270	0.827
			Load	1.000	1.000	1.001	<0.001	39	0.063
			Load'	0.999	0.999	1.000	<0.001	21	0.143
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.053	0.645	1.719	0.25	11521	0.837
			Age (Years)	0.87	0.67	1.129	0.133	1146	0.294
	Daily ACWR 7:21-period	Same day	Intercept	0.041	0.001	2.833	2.157	3372	0.140
			Load	0.743	0.362	1.523	0.366	1301	0.417
			Load'	1.648	0.687	3.952	0.445	394	0.262
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.127	0.729	1.741	0.222	8737	0.591
			Age (Years)	0.989	0.776	1.259	0.124	3357	0.926
		Next 4 days	Intercept	0.234	0.001	99.719	3.088	2022	0.638
			Load	2.006	1.006	4.002	0.348	98	0.048
			Load'	0.292	0.133	0.643	0.395	70	0.003
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.316	0.708	2.449	0.317	7490	0.385
			Age (Years)	0.886	0.624	1.257	0.179	1426	0.497

Continues on the next page

Population	Load Definition[1]	Injury Definition[2]	Variable	OR[3]	CI 2.50 %	CI 97.50 %	SE	df	p
Handball	Micro-cycle ACWR 1:3-period	Next micro-cycle	Intercept	0.165	0.003	9.425	2.062	1450	0.382
			Load	0.878	0.397	1.939	0.404	955	0.747
			Load'	1.335	0.599	2.976	0.408	969	0.479
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	0.908	0.596	1.384	0.215	1551	0.654
			Age (Years)	1.004	0.795	1.267	0.119	1313	0.976

Abbreviations: ACWR, Acute: Chronic Workload Ratio; CI, 95% Confidence Intervals; df, Degrees of Freedom; Football Elite, Norwegian Premier League; OR, Odds Ratio; SE, Standard Error; sRPE, daily session Rating of Perceived Exertion.
[1]Daily ACWR 7:21-period was the 7-day acute sRPE divided by previous 21-day chronic sRPE per day; Micro-cycle ACWR 1:3-period was the 1-micro-cycle acute sRPE divided by previous 3-micro-cycle chronic sRPE per micro-cycle. A micro-cycle was defined as all recovery days after the previous match as well as the training days before the next match.
[2]Same day was injury same day as the measured load value; Next 4 days was one or more injuries during the four days after the measured load value; Next micro-cycle was one or more injuries during the micro-cycle after the micro-cycle of the measured load values.
[3]As load was fitted with cubic splines, the effect-size, Odds Ratio, is uninterpretable for this parameter.

Table S4. Odds ratio with 95% confidence intervals, standard error, degrees of freedom and p-values from modelling the relationship between training load and injury risk using mixed effect logistic regression models which assume linearity.

Population	Load Definition[1]	Injury Definition[2]	Variable	OR	CI 2.5%	CI 97.5%	SE	df	p
Football U-19	sRPE	Same day	Intercept	0.003	<0.001	0.795	2.781	942	0.041
			Load	1.000	0.999	1.001	<0.001	3338	0.755
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.165	0.629	2.156	0.314	3287	0.627
			Age (Years)	1.088	0.800	1.478	0.156	922	0.591
		Next 4 days	Intercept	0.034	<0.001	15.046	3.099	263	0.275
			Load	1.000	1.000	1.001	<0.001	3015	0.067
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.186	0.585	2.406	0.361	4759	0.636
			Age (Years)	0.969	0.686	1.369	0.175	256	0.86
	Daily ACWR 7:21-period	Same day	Intercept	0.001	<0.001	0.405	3.181	1341	0.025
			Load	1.346	0.859	2.107	0.228	492	0.194
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.329	0.645	2.737	0.368	2601	0.440
			Age (Years)	1.139	0.798	1.625	0.181	1289	0.473
		Next 4 days	Intercept	0.003	<0.001	98.996	5.313	108	0.266
			Load	1.312	0.633	2.722	0.372	1744	0.465
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.408	0.478	4.147	0.551	2514	0.535
			Age (Years)	1.052	0.583	1.899	0.298	108	0.865

Continues on the next page

Population	Load Definition[1]	Injury Definition[2]	Variable	OR	CI 2.5%	CI 97.5%	SE	df	p
Football U-19	Micro-cycle ACWR 1:3-period	Next micro-cycle	Intercept	0.013	<0.001	10.059	3.396	331	0.199
			Load	0.850	0.324	2.232	0.492	534	0.741
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.088	0.475	2.492	0.422	645	0.842
			Age (Years)	1.079	0.738	1.577	0.193	299	0.694
Football Elite	sRPE	Same day	Intercept	0.001	<0.001	0.008	1.344	4481	<0.001
			Load	1.000	0.999	1.002	0.001	4481	0.867
			Age (Years)	1.096	0.998	1.204	0.048	4481	0.055
		Next 4 days	Intercept	<0.001	<0.001	0.014	2.572	1412	<0.001
			Load	1.001	0.998	1.003	0.001	15	0.484
			Age (Years)	1.189	0.994	1.422	0.091	1662	0.058
	Daily ACWR 7:21-period	Same day	Intercept	<0.001	<0.001	0.008	1.555	3576	<0.001
			Load	1.255	0.459	3.427	0.512	552	0.658
			Age (Years)	1.102	0.994	1.222	0.052	3847	0.064
		Next 4 days	Intercept	<0.001	<0.001	0.042	2.952	1288	0.002
			Load	0.739	0.253	2.165	0.539	70	0.577
			Age (Years)	1.189	0.974	1.452	0.102	1356	0.089
	Micro-cycle ACWR 1:3-period	Next micro-cycle	Intercept	0.001	<0.001	0.036	1.665	186	<0.001
			Load	2.183	0.350	13.625	0.910	47	0.396
			Age (Years)	1.115	1.018	1.221	0.046	476	0.019
Handball	sRPE	Same day	Intercept	0.063	0.002	2.082	1.782	1673	0.121
			Load	1.000	1.000	1.000	0.000	7341	0.240
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.124	0.788	1.604	0.181	11869	0.519
			Age (Years)	0.953	0.779	1.166	0.103	1739	0.638

Continues on the next page

Population	Load Definition[1]	Injury Definition[2]	Variable	OR	CI 2.5%	CI 97.5%	SE	df	p
Handball	sRPE	Next 4 days	Intercept	0.603	0.006	56.793	2.317	1367	0.827
			Load	1.000	1.000	1.000	<0.001	64	0.348
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.045	0.641	1.705	0.250	11524	0.859
			Age (Years)	0.873	0.671	1.135	0.134	1420	0.310
	Daily ACWR 7:21-period	Same day	Intercept	0.030	<0.001	2.084	2.159	2373	0.105
			Load	1.106	0.846	1.445	0.136	204	0.459
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.129	0.730	1.748	0.223	8738	0.585
			Age (Years)	0.988	0.775	1.260	0.124	3373	0.921
		Next 4 days	Intercept	0.535	0.001	233.805	3.098	899	0.840
			Load	0.895	0.599	1.338	0.203	118	0.587
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	1.319	0.710	2.452	0.316	7582	0.381
			Age (Years)	0.874	0.615	1.243	0.179	796	0.454
	Micro-cycle ACWR 1:3-period	Next micro-cycle	Intercept	0.141	0.002	12.674	2.293	1448	0.393
			Load	1.125	0.716	1.769	0.230	771	0.609
			Sex Female (Ref)	-	-	-	-	-	-
			Sex Male	0.908	0.567	1.453	0.240	1552	0.686
			Age (Years)	1.002	0.773	1.300	0.133	1404	0.986

Abbreviations: ACWR, Acute: Chronic Workload Ratio; CI, 95% Confidence Intervals; df, Degrees of Freedom; Football Elite, Norwegian Premier League; OR, Odds Ratio; SE, Standard Error; sRPE, daily session Rating of Perceived Exertion.

[1]Daily ACWR 7:21-period was the 7-day acute sRPE divided by previous 21-day chronic sRPE per day; Micro-cycle ACWR 1:3-period was the 1-micro-cycle acute sRPE divided by previous 3-micro-cycle chronic sRPE per micro-cycle. A micro-cycle was defined as all recovery days after the previous match as well as the training days before the next match.

[2]Same day was injury same day as the measured load value; Next 4 days was one or more injuries during the four days after the measured load value; Next micro-cycle was one or more injuries during the micro-cycle after the micro-cycle of the measured load values.

Figures

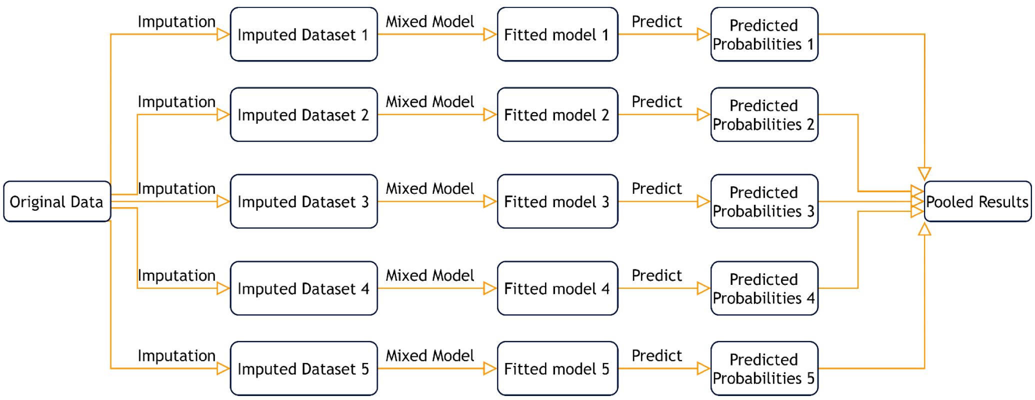

Figure S1. Illustration of the modelling process in the framework of multiple imputation. Following the recommendations in "Flexible Imputation of Missing Data, Second Edition" by Stef van Buuren,[1] which is also available online.[2] Missing load and age values were predicted and imputed using predictive mean matching.[3] 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,[4] but was not itself imputed before analysis (guides in Van Buuren [1] 6.3.2, 6.4.1).[5] The number of imputed datasets, five, is recommended in most cases (Van Buuren section 2.8).[1] A mixed logistic regression model was run on each dataset, returning five fitted models. Each model was used to make predictions, and the mean of the predicted probabilities was used in final visualization, then the model parameters were pooled using Rubin's rules.[6]

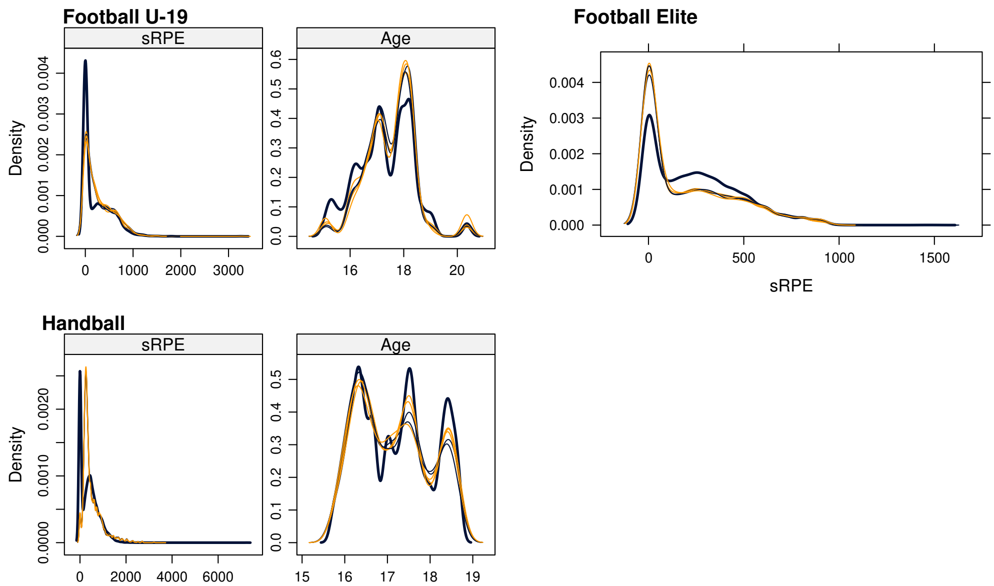

Figure S2. Distribution of real data values (blue) compared to imputed values from five imputed datasets (yellow) for the session Rating of Perceived Exertion (sRPE) measured in arbitrary units, and Age (years) in the Norwegian elite U-19 football dataset (Football U-19), the Norwegian Premier League dataset (Football Elite), and the Norwegian elite youth handball dataset. The Norwegian Premier League dataset had no missing age values.

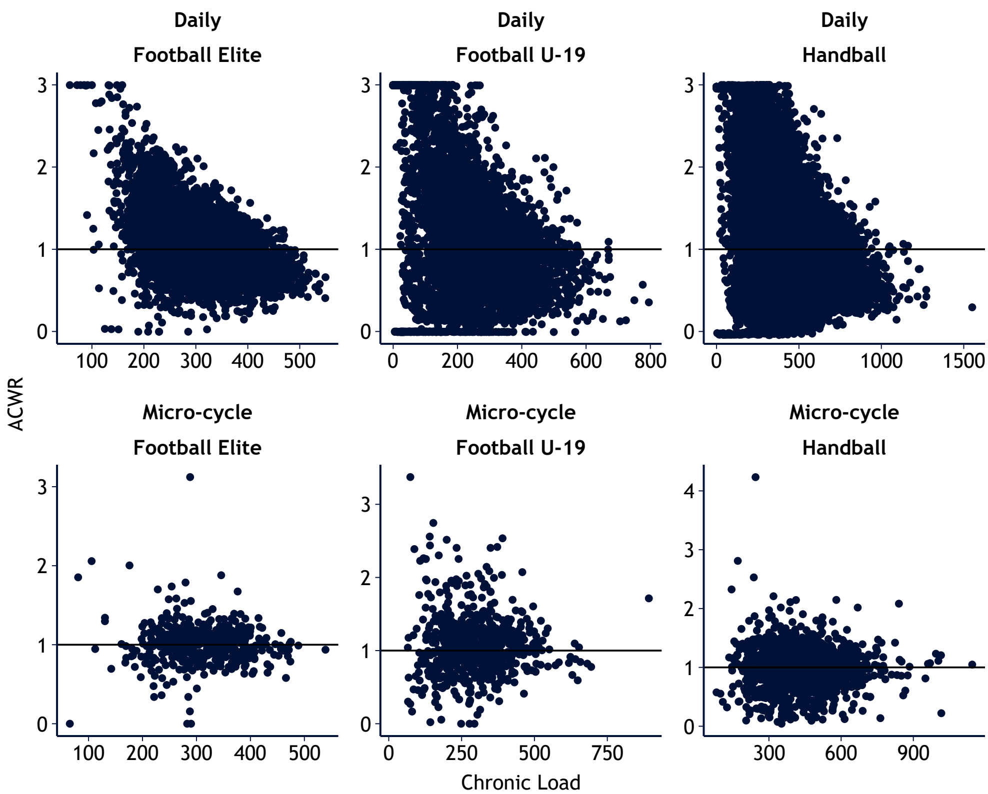

Figure S3. Scatterplot of Acute:Chronic Workload Ratio (ACWR) value vs. corresponding chronic load value (the denominator) in the Norwegian Premier League football dataset (Football Elite), the Norwegian elite U-19 football dataset (Football U-19), and Norwegian elite youth handball dataset (Handball). When computing a ratio, one assumes that there is no relationship between the ratio and the denominator after controlling for the denominator; a ratio is only effective when the relationship between the numerator and the denominator is a straight line that intersects the origin.[7] For micro-cycle ACWR, the assumption is upheld, while for daily ACWR, the assumption is violated.

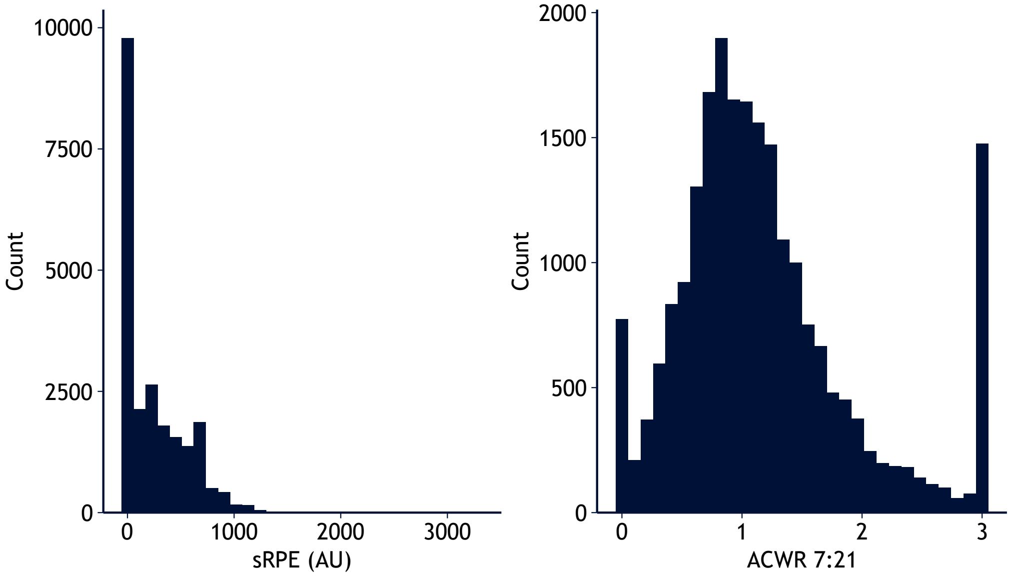

Figure S4. Distribution of the session Rating of Perceived Exertion (sRPE) reported in arbitrary units (AU), and distribution of the 7-day Acute Workload divided by 21-Chronic Workload (ACWR 7:21), from the Norwegian elite U-19 football data used as basis for simulations.

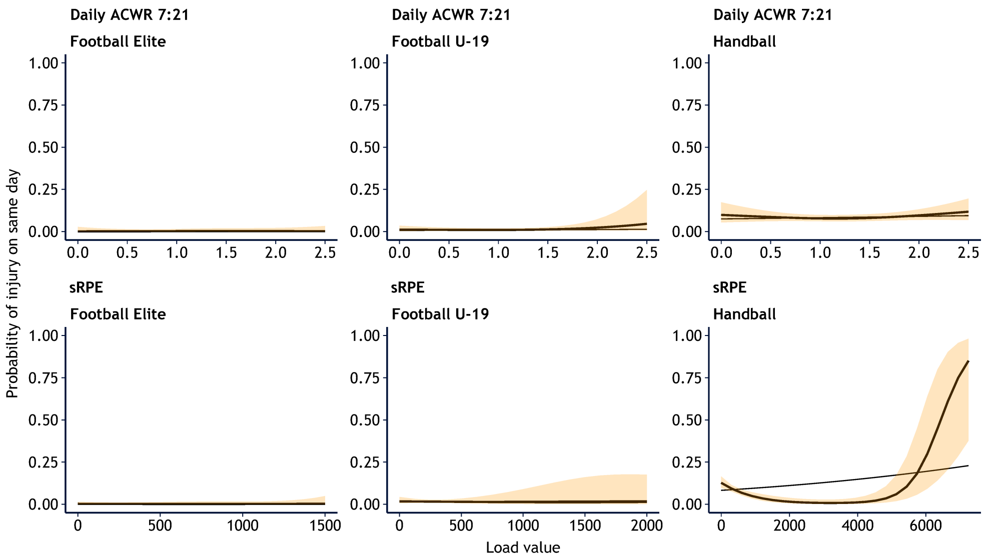

Figure S5. Probability of injury on the same day for each level of session Rating of Perceived Exertion (sRPE) and level of daily Acute:Chronic Workload Ratio (ACWR), in Norwegian Premier League (Football Elite), Norwegian elite U-19 football (Football U-19), and Norwegian elite youth handball (Handball). Probabilities are predicted by mixed-effects logistic regression models with restricted cubic splines. The yellow area represents 95% confidence intervals around predicted values. The straight line shows the same predictions from an equivalent model without splines (i.e. assuming linearity).

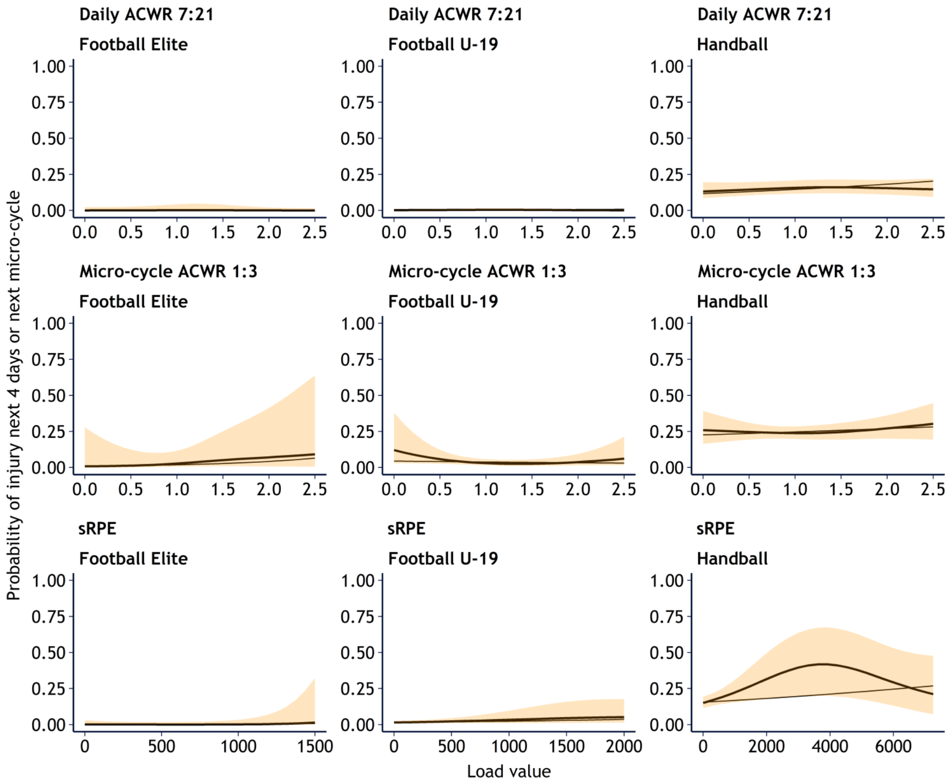

Figure S6. Probability of injury in the future for each level of daily Acute:Chronic Workload Ratio (ACWR), level of Micro-cycle ACWR, and level of session Rating of Perceived Exertion (sRPE), in Norwegian Premier League (Football Elite), Norwegian elite U-19 football (Football U-19), and Norwegian elite youth handball (Handball). Future injury was defined as any injury occurring during the next 4 days for all models except micro-cycle models, where future injury was defined as any injury occurring during the next micro-cycle. Probabilities are predicted by mixed-effects logistic regression models with restricted cubic splines. The yellow area represents 95% confidence intervals around predicted values. The straight line shows the same predictions from an equivalent model without splines (i.e. assuming linearity).

REFERENCES

1. Van Buuren S. Flexible imputation of missing data: Chapman and Hall/CRC 2018.
2. Van Buuren S. Flexible Imputation of Missing Data: CRC Press; 2018 [Second Edition:[Available from: https://stefvanbuuren.name/fimd/ accessed 2021-02-12.
3. Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *American journal of epidemiology* 2004;160(1):34-45.
4. 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.
5. 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.
6. Miles A. Obtaining predictions from models fit to multiply imputed data. *Sociological Methods & Research* 2016;45(1):175-85.
7. Curran-Everett D. Explorations in statistics: the analysis of ratios and normalized data. *Advances in physiology education* 2013;37(3):213-19.

SUPPLEMENTARY MATERIALS AND METHODS

Participants

To find out whether the relationship between training load and injury risk may be non-linear, and whether the shape may vary between different populations, access was gained to data from different sports: football (soccer) and handball, and different populations within the same sport: Norwegian elite U-19 football data and a Norwegian Premier League football team.

The Norwegian elite U-19 data was used in Dalen-Lorentsen, et al. [1]. It was a cohort of six Norwegian elite U-19 football teams (3 female and 3 male) with 81 players (55% male, mean age: 17 years, standard deviation (SD): 1 year) followed from July to October 2017 for 104 days.

The second football cohort was a professional male football team from the Norwegian Premier League surveyed from January to December 2019 for 323 days (n = 36, mean age: 26 years (SD: 4)).[2]

The handball 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.[3]

Training load definition

In all three cohorts, players reported the number of training sessions and matches daily. They also reported the duration of each activity and their Rating of Perceived Exertion (RPE)[4] on the modified Borg CR10 scale.[5] To derive the session RPE (sRPE),[5] we multiplied the RPE by the activity duration in minutes. To summarize daily loads, sRPE was calculated for each session and subsequently summed.

Missing sRPE values are reported in Table S1 (Supplementary I) and were 24% for elite U-19 football, 41% for Premier League football, and 64% for elite youth handball. The values were imputed using multiple imputation, a method that also performs well in cases of high amounts of missing (80%) if the data are Missing at Random,[6] which is most common in clinical research.[7] For more detailed information on the imputation process, see Supplementary I Figure S1. The observed distribution was maintained in the imputed values; therefore the imputation was deemed valid (Figure S2).

All load measures were based on players' daily ratings of perceived exertion (sRPE). We calculated an Acute-Chronic Workload Ratio (ACWR) in two different ways:

Daily ACWR 7:21

The mean sRPE across 7 days divided by the exponentially-weighted-moving average (EWMA) of the previous 21 days (Figure 1). EWMA accounts for the assumption that load values closer in time to the event are more associated with the event than measures further back in time.[8] The calculation was uncoupled, meaning that the 7 days of acute load for the numerator were not included in the 21 days of the denominator.[9]

Bache-Mathiesen LK, *et al. BMJ Open Sp Ex Med* 2021; 7:e001119. doi: 10.1136/bmjsem-2021-001119

The calculation was performed on a sliding window moving one day at a time from and including the 28th day.[10] The last day in the acute load is considered Day 0 (Figure 1).

One limitation with the ACWR is that it bloats cases where the athlete has had little to no chronic load and returns to regular exercise. In previous studies, these cases have traditionally been deleted.[11] Here, these cases were set to have an ACWR of 3, a very high ACWR value, in line with recommendations in Harrell [12] for treatment of overly influential values. Likewise, if the EWMA chronic load was equal to zero and ACWR could not be calculated, the ACWR was set to 3.

Micro-cycle ACWR 1:3

The mean sRPE for each micro-cycle divided by the EWMA of the previous 3 micro-cycles, uncoupled (Figure 1). A micro-cycle was defined as all recovery days after the previous match and the training days before the next match. The next micro-cycle started on the first training day after the match, and so on. For an illustration of a micro-cycle, see Figure 1. The calculation was performed in the same manner as described for daily ACWR, on a sliding window moving one micro-cycle at a time from and including the 4th micro-cycle. The last day of the 4th micro-cycle was considered Day 0 (Figure 1).

When computing a ratio, one assumes that there is no relationship between the ratio and the denominator after controlling for the denominator; a ratio is only effective when the relationship between the numerator and the denominator is a straight line that intersects the origin.[13] Tests of this assumption are reported in Supplementary I Figure S3.

Injury definition

The same online questionnaire was used to collect daily health status and training information from all three sports cohorts. The elite U-19 football data and elite youth handball data were collected via the Briteback AB online survey platform, while the Norwegian Premier League football data were collected with Athlete Monitoring, Moncton, Canada.

The players daily reported whether they had experienced “no health problem”, “a new health problem”, or an “exacerbation of an existing health problem”. In the youth elite handball study, if players reported any new health problems, they were immediately prompted to specify whether it was an injury or illness in the questionnaire. In the football studies, if players reported any new health problems, a clinician contacted them by telephone the following day for a structured interview and classified the health problem as an injury or illness with the Union of European Football Associations guidelines.[14] Players were asked to report all physical complaints, irrespective of their consequences on sports participation or the need to seek medical attention.[15]

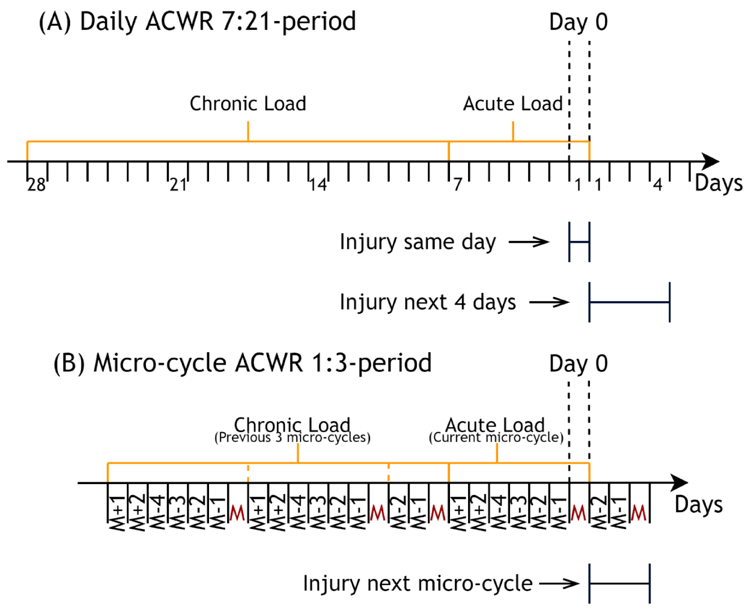

Figure 1. Illustration of time-periods for calculating (A) Daily ACWR 7:21-period and (B) Micro-cycle ACWR 1:3-period. The first day that ACWR is calculated from is denoted Day 0. The space between two tick marks represent one day (24 hours). For (B), a micro-cycle period consists of all activity before a new match (M). That is, recovery days after the previous match as well as the training days before the next match. Days denoted with negative numbers are training days before the next match (M-1; being the day before the match, M-2; two days before a match, and so on). Days with positive numbers are recovery and training days after a match (M+1; being the day after a match, M+2; two days after a match). The number of days between matches varies by the match schedule. How a team plan their training and recovery activities varies, and is dependent on the teams' philosophy. For (A), injury on the same day is defined as an injury on Day 0, and future injury is defined as an injury occurring during the next 4 days excluding Day 0. For (B) future injury was defined as an injury occurring during the next micro-cycle excluding Day 0.

Ethical Considerations

Data collection for all three 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 elite U-19 football (5487); Norwegian Premier League football (722773); Norwegian elite youth handball (407930). The Norwegian elite U-19 football study was also approved by the South-Eastern Norway Regional Committee for Medical and Health Research Ethics (2017/1015). Ethical principles were followed in accordance with the Declaration of Helsinki.[16] All participants provided informed consent. All participants were above the age of 15 and parental consent was not required. Participants were assured their responses would only be available to the research team, participation was voluntary, and consent could be withdrawn at any time.

Legality of using the data in this study was dependent on the "purposes of the processing for which the personal data were intended" as written in the consent forms.[17] The consent forms for the football studies were general enough that use in this study were within the posted aims. For the elite youth handball data, the Norwegian Centre for Research Data deemed the aims described in the consent forms invalid for use in this study, and the data had to be anonymised. Anonymisation was performed under guidelines outlined by The Norwegian Data Protection Authority.[18]

Statistical analyses

To estimate the relationship between training load and injury risk, mixed-effects logistic regression was used. Logistic regression is the most frequent regression analysis in the field of training load and injury.[19] Mixed models have been recommended to account for within-player dependencies[20] and are robust to missing data in the outcome variable.[21]

All injuries were considered an event in the response variable. Illnesses and explicit replies of "no health problem" were considered non-events. Non-responses were recorded as missing. Independence between subsequent injuries within the same player was assumed.

We considered two outcomes: (1) occurrence of an injury on the same day as the observed training load (Day 0); (2) occurrence of injury in the future, where the current observation day (Day 0) was not included. For unmodified training load values and daily ACWR 7:21-period, future injury was defined as an injury occurring during the next four days excluding Day 0. For micro-cycle ACWR 1:3-period, the future injury was any injury occurring during the next micro-cycle excluding Day 0. See Figure 1 for an illustration of injury time periods and Table S2 (Supplementary I) for a list of the different models.

For models where the injury definition was set to the future, any number of injuries sustained during the time window were aggregated to 1 event. Furthermore, injuries sustained before the first calculated ACWR value had to be discarded. Consequentially, the number of injuries included in the different models varied (Table S2).

We adjusted for player age in all analyses. In addition, we adjusted for sex in the U-19 elite football and the elite youth handball models. Akaike's Information Criterion (AIC) was used to determine the model fit between including a random intercept only vs. including a random intercept & random slope for training load per player, where the best fit was chosen for the final model. Overly influential observations – extreme outliers which affect analyses – were checked using dfbeta.[12]

In all models, the relationship between sRPE and injury risk was modelled with Restricted Cubic Splines (RCS).[22] The number of knots was decided using AIC. The models were repeated without splines to simulate the relationship we would have discovered if we had assumed linearity. When using RCS, the estimated regression coefficients do not have a clinically meaningful interpretation, and only their p-values are numerically interpretable.[12] The main result is therefore a visualization of the model predictions (with 95% cluster-robust confidence intervals) to determine the shape of the relationship between training load and injury risk. To limit the number of figures to the most relevant, only predictions from models that showed a tendency towards a relationship or stronger are included in the article itself, but figures for all relationships are shown in Supplementary I Figure S5–S6. For each model, predicted values were estimated on each imputed dataset, and then pooled before visualization (Figure S1).[23]

Our analyses served to illustrate whether there is any evidence for non-linearity in training load and injury research and should not be interpreted as causal inference.

Simulation

Step 1 Preparing data

In addition to analysing real data, we performed (stochastic) simulations to compare different methods for ascertaining non-linear and linear relationships between training load and injury risk. The methodology here is focused on a causal research setting; however, the methods may also be applied in predictive research.[25] The simulations were based on the elite U-19 football dataset since it had the least missing data (24%). An imputed dataset was chosen from the 5 datasets previously imputed with multiple imputation.

Two datasets were created. The first kept the original 8 495 sRPE and 6 308 ACWR values. In the second, sRPE and ACWR were sampled with replacement to generate a scenario of 3 football teams (75 players) followed meticulously for a season (300 days), altogether 22 500 training load values. The distribution of the real data was retained during sampling; highly skewed for sRPE and Gaussian for ACWR (Figure S4).

Step 2 Generating predetermined relationships

Artificial injuries were simulated and added to each dataset under different relationship scenarios with training load. The risk models were based on the logistic function:

$$logistic(x) = \frac{1}{1 + \exp(-x)}$$

U shape

A symmetrical U parabola coinciding with the theory in Gamble 2013.[24] Using the logistic function above, the U shape function was:

$$Prob\{Y = 1|sRPE = \ logistic(-1 + \ 0.0000002 * (sRPE - 1500)^2)$$

Where Y is an indicator variable for injury.

J shape

The J shape was chosen to reproduce findings in Carey, et al. [25] with the risk function:

$$Prob\{Y = 1|ACWR\} = logistic(\begin{cases} -3.4 + 2 \cdot (1 - ACWR)^2, & ACWR < 1 \\ -3.4 + (1 - ACWR)^2, & 1 \leq ACWR < 1.7 \\ 1.5 \cdot ACWR - 5.4, & ACWR \geq 1.7 \end{cases})$$

Linear shape

A linear shape to determine whether a method optimal for non-linear modeling can also model a linear shape. The function was then:

$$Prob\{Y = 1|sRPE\} = logistic(-0.5 + 0.001 * sRPE)$$

For the U shape and linear shape, the simulated probability of an injury was based on the sRPE, while for the J shape, it was based on the ACWR.

We assumed a longitudinal design for the simulation, and an autoregressive correlation structure was implemented to ensure that values closer in time were more highly correlated than values further apart.[8] Any reference to the "true" probability refers to the simulated probability we have created for a given scenario, and which we aim to model.

While shown to be valid and reliable, the sRPE may still have some measurement error.[26] Before analyses, noise was added to load values to simulate this. The amount was set to the default jitter value, which was:

$$\frac{\max(load) - \min(load)}{50}$$

Step 3 Running models on all combinations of datasets and relationship shapes

In the same manner as in the analysis of the real data, a logistic regression model with random effects (mixed model) was used to determine the relationship between training load and predefined injury risk. Different methods of modifying training load were compared.

Linear Model

A standard logistic regression served as an example of a method which assumes linearity and illustrated the degree of error should the linearity assumption be ignored in cases where the relationship is non-linear. The purpose was to determine whether more complicated or time-consuming methods were worth the effort.

A logistic regression model describes the relationship between the probability of an event in the response variable Y (injury), given the status of the explanatory variables $X = \{x_1, x_2, \dots, x_n\}$ as the additive contribution of the intercept β_0 and linear slopes $\beta_1, \beta_2, \dots, \beta_n$ of said variables.[27] In a logistic regression with a single explanatory variable (covariate) x_1, representing the load variable, the formula is as follows:

$$Prob\{Y = 1|X\} = \frac{\exp(\beta_0 + \beta_1 x_1 + \gamma)}{1 + \exp(\beta_0 + \beta_1 x_1 + \gamma)} = logistic(\beta_0 + \beta_1 x_1 + \gamma)$$

Where γ is the random effect term.

Categorization

Although categorizing the load variable into groups before performing the intended analysis has previously been shown to be a poor method for modelling non-linear relationships,[25] we chose nevertheless to include it in our comparison of methods. For one, the method has been recommended since.[28 29] For another, as the authors requested, we attempted to reproduce the results in another sport population under different conditions. Here, the sRPE data are highly skewed. We also increased the number of permutations for more accurate results.

To show how results may differ depending on how variables are categorized, we categorized the training load variable in two ways, before including them in two separate logistic regression models. The first was a categorization by quartiles to exemplify a data-driven approach, a chosen method in numerous studies in the past.[30-32] The second was subjectively chosen cut-offs based on the range of the data. For sRPE, four categories were made: <= 499, 500–1 499, 1 500–2 499 and >= 2 500. For ACWR, three categories were made: < 1, 1–1.74 and >= 1.75, which are the same used in Carey, et al. [25].

Quadratic model
Quadratic regression has seen some use in recent years.[33] In some studies, a quadratic term was added to the regression model to test for linearity.[34 35] Where as in others, the researchers hypothesized a parabolic shape and used quadratic regression to model the training load and injury relationship accordingly.[10 36] In a quadratic model, a polynomial to the second power is added to the standard regression model. For the logistic regression, it is denoted thus:

$$Prob\{Y = 1|X\} = logistic(\beta_0 + \beta_1 x_1 + \beta_2 {x_1}^2 + \gamma)$$

The model will then fit a parabolic shape between the probability of an event in Y (injury) and the explanatory variable x_1 (training load). A polynomial term can be added regardless of whether it is a linear, logistic or Poisson regression model. Although easy-to-use and intuitive, the main disadvantage of quadratic regression is that it can only model a parabola; for instance, it cannot uncover a sigmoidal shape.

Fractional polynomials
Quadratic regression is a sub-method of the more flexible Fractional Polynomials (FP), which has been used in one single training load and injury risk study.[37] Fractional polynomials, simply put, uses polynomial transformations to estimate the association between the covariate and the outcome.[38] FPs can model multiple shapes, not just the parabola. Fractional polynomials add either a single polynomial term to the pth power to the regression model (known as an FP1 model), or two polynomial terms to the pth power to the model (FP2 model).[38] The FP2 model has been shown to be the optimal choice in most cases and was chosen for all models in this study.[39] The logistic regression model with FP2 is as follows:

$$Prob\{Y = 1|X\} = logistic(\beta_0 + \beta_1 x_1 + \beta_2 {x_1}^{p1} + \beta_3 {x_1}^{p2} + \gamma)$$

Where $p1$ and $p2$ are exponents selected from {-2, -1, -0.5, 0, 0.5, 1, 2, 3}. A form of backward elimination was used to determine the polynomial powers with the best fit, see Ambler and Benner [40] for more details. A step-by-step guide to perform FP in R can be accessed on the primary author's GitHub.[41]

Restricted cubic splines
Another possible approach to model non-linear relationships is to use Restricted Cubic Splines (RCS). This approach as well as FP, performed better than categorization in the study by Carey, et al. [25], who found no distinct differences between RCS and FP. In cubic splines, the X-axis is divided into intervals by a number of endpoints (knots). At these knots, different cubic polynomials are joined and forced to have a consistent function, slope and acceleration (second derivative) until the next knot. At the knot, the rate change of acceleration (third derivative) may change. For three knots a, b and c, our logistic regression formula becomes:

$$Prob\{Y = 1|X\} =$$
$$logistic[\beta_0 + \beta_1 x_1 + \beta_2 {x_1}^2 + \beta_3 {x_1}^3 + \beta_4 (x_1 - a)^3 + \beta_5 (x_1 - b)^3 + \beta_6 (x_1 - c)^3 + \gamma]$$

In restricted cubic splines, the function is restricted to behave linearly in the tails.[22]

RCS has the advantage of flexibility, but the effect sizes are difficult to interpret, and the number and location of knots must be chosen, either by a data-driven or approach or as a choice of the user. As 3–5 knots are appropriate for most datasets,[12] 3 knots were used in all simulation models. We compared two different ways of choosing knot location. In the first, the knot locations were chosen by the default approach in the statistical software (data-driven), and in the other, knot locations were cut-off subjectively at sRPE = 500, 1 500 and 2 500, and likewise at ACWR = 1, 1.75 and 2, to cover the range of the load metrics.

A step-by-step guide to perform RCS in R can be accessed on the primary author's GitHub.[42]

Step 4 Calculating performance metrics

The Root-Mean-Squared Error (RMSE) was calculated to numerically evaluate the accuracy of the methods. RMSE is a combined measure of accuracy and precision, where the lower the RMSE, the better the method. RMSE was calculated as the square root of the mean difference between the true risk and predicted risk for each observation. The scale of the RMSE depends on the analysis in question, and it is therefore only interpretable by comparing values in the same analysis – the values cannot be interpreted in isolation.[43]

To supplement RMSE, the proportion of prediction intervals that included the true coefficient was calculated (coverage). Brier score for model fit and C-statistics (also known as the concordance, or as the area under the receiving operating characteristic curve) was calculated for predictive ability, since they are commonly used in training load and injury risk studies.[44-47]

Final analyses

In summary, the four steps of the simulation were:

1 Sample training load values from the elite U-19 football data
2 Simulate injuries with three different shapes for the relationship between injury risk and training load

3 Fit seven different models with injury as the outcome and training load as the explanatory variable
4 Calculate performance measures

Using formulas listen in Morris, et al. [43], accepting a Monte Carlo Standard Error of no more than 0.5, the number of permutations needed for an accurate determination of coverage was:

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

Steps 1–4 were therefore repeated 1 900 times for all relationship scenarios.

For the U-shaped relationship, predicted values were visualized alongside the predefined shape to determine each method's ability to capture the true relationship. Only one permutation was used for the visualization to avoid cluttering of lines.

The mean RMSE, coverage, C-statistics and Brier score were calculated for each combination of model-method and dataset sizes for the U-, J- and linear-shaped relationships. As mean RMSE was the most relevant metric for determining model accuracy, it was visually compared for the non-linear shapes.

All statistical analyses and simulations were performed using R version 4.0.2[48] with RStudio version 1.3.1056. Packages were used for specific purposes: multiple imputation with MICE,[49] mixed models with lme4,[50] predictions with ggeffects,[51] confidence intervals with clubSandwich,[52] predictions with prediction intervals using merTools,[53] and splines with the rms package.[54] The simulations were run on a computer with an Intel(R) Core(TM) i7-6700K 4.00GHz CPU, and with 16 GB RAM. A GitHub repository is available with all R code and the data used in the simulations.[55]

REFERENCES

1. 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.
2. Theron GF. The use of Data Mining for Predicting Injuries in Professional Football Players. University of Oslo, 2020.
3. 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
4. Borg G, Hassmén P, Lagerström M. Perceived exertion related to heart rate and blood lactate during arm and leg exercise. *European journal of applied physiology and occupational physiology* 1987;56(6):679-85.
5. 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.
6. Madley-Dowd P, Hughes R, Tilling K, et al. The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of clinical epidemiology* 2019;110:63-73.

7. Janssen KJ, Donders ART, Harrell Jr FE, et al. Missing covariate data in medical research: to impute is better than to ignore. *Journal of clinical epidemiology* 2010;63(7):721-27.
8. 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.
9. 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.
10. 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.
11. Hulin BT, Gabbett TJ, Lawson DW, et al. The acute: chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *Br J Sports Med* 2016;50(4):231-36.
12. Harrell FE. Regression modeling strategies. *BIOS* 2017;330:2018.
13. Curran-Everett D. Explorations in statistics: the analysis of ratios and normalized data. *Advances in physiology education* 2013;37(3):213-19.
14. Hägglund M, Waldén M, Bahr R, et al. Methods for epidemiological study of injuries to professional football players: developing the UEFA model. *British journal of sports medicine* 2005;39(6):340-46.
15. 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.
16. Malik AY, Foster C. The revised Declaration of Helsinki: cosmetic or real change? *Journal of the Royal Society of Medicine* 2016;109(5):184-89.
17. EU P. General Data Protection Agency Article 13: intersoft consulting; 2018 [Available from: https://gdpr-info.eu/art-13-gdpr/ accessed 2021-02-10.
18. 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 2021-04-12.
19. 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.
20. Nielsen RO, Shrier I, Casals M, et al. Statement on Methods in Sport Injury Research From the First METHODS MATTER Meeting, Copenhagen, 2019. *Journal of Orthopaedic & Sports Physical Therapy* 2020;50(5):226-33.
21. 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.
22. Durrleman S, Simon R. Flexible regression models with cubic splines. *Statistics in medicine* 1989;8(5):551-61.
23. Miles A. Obtaining predictions from models fit to multiply imputed data. *Sociological Methods & Research* 2016;45(1):175-85.
24. Gamble P. Reducing injury in elite sport-is simply restricting workloads really the answer? *New Zealand Journal of Sports Medicine* 2013:36.
25. Carey DL, Crossley KM, Whiteley R, et al. Modeling Training Loads and Injuries: The Dangers of Discretization. *Medicine and science in sports and exercise* 2018;50(11):2267-76.
26. Haddad M, Stylianides G, Djaoui L, et al. Session-RPE method for training load monitoring: validity, ecological usefulness, and influencing factors. *Frontiers in neuroscience* 2017;11:612.
27. Silk MJ, Harrison XA, Hodgson DJ. Perils and pitfalls of mixed-effects regression models in biology. *PeerJ* 2020;8:e9522.
28. Ruddy JD, Cormack SJ, Whiteley R, et al. Modeling the risk of team sport injuries: a narrative review of different statistical approaches. *Frontiers in physiology* 2019;10

29. 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.
30. Cross MJ, Williams S, Trewartha G, et al. The influence of in-season training loads on injury risk in professional rugby union. *International journal of sports physiology and performance* 2016;11(3):350-55.
31. Malone S, Hughes B, Doran DA, et al. Can the workload–injury relationship be moderated by improved strength, speed and repeated-sprint qualities? *Journal of science and medicine in sport* 2019;22(1):29-34.
32. Stares J, Dawson B, Peeling P, et al. Identifying high risk loading conditions for in-season injury in elite Australian football players. *Journal of science and medicine in sport* 2018;21(1):46-51.
33. Gabbett TJ. The training—injury prevention paradox: should athletes be training smarter and harder? *British journal of sports medicine* 2016;50(5):273-80.
34. Sampson JA, Murray A, Williams S, et al. Injury risk-workload associations in NCAA American college football. *Journal of science and medicine in sport* 2018;21(12):1215-20.
35. Warren A, Williams S, McCaig S, et al. High acute: chronic workloads are associated with injury in England & Wales Cricket Board Development Programme fast bowlers. *Journal of science and medicine in sport* 2018;21(1):40-45.
36. 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.
37. Lathlean TJ, Gastin PB, Newstead SV, et al. Absolute and Relative Load and Injury in Elite Junior Australian Football Players Over 1 Season. *International Journal of Sports Physiology and Performance* 2019;1(aop):1-9.
38. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 1994;43(3):429-53.
39. Binder H, Sauerbrei W, Royston P. Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. *Statistics in Medicine* 2013;32(13):2262-77.
40. mfp: Multivariable Fractional Polynomials. R package version [program], 2015.
41. Bache-Mathiesen L. Performing Fractional Polynomials in R [PDF]. GitHub; 2021 [Available from: https://github.com/lenakba/load-injury-non-linearity-study/blob/main/guide_fractional_polynomials.pdf accessed 2021-04-12.
42. Bache-Mathiesen L. Performing Restricted Cubic Splines in R [PDF]. GitHub; 2021 [Available from: https://github.com/lenakba/load-injury-non-linearity-study/blob/main/guide_restricted_cubic_splines.pdf accessed 2021-04-12.
43. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in medicine* 2019;38(11):2074-102.
44. Johnston R, Cahalan R, Bonnett L, et al. Training load and baseline characteristics associated with new injury/pain within an endurance sporting population: a prospective study. *International journal of sports physiology and performance* 2019;14(5):590-97.
45. McCall A, Dupont G, Ekstrand J. Internal workload and non-contact injury: a one-season study of five teams from the UEFA Elite Club Injury Study. *British journal of sports medicine* 2018;52(23):1517-22.
46. Thornton HR, Delaney JA, Duthie GM, et al. Importance of various training-load measures in injury incidence of professional rugby league athletes. *International journal of sports physiology and performance* 2017;12(6):819-24.
47. Visnes H, Bahr R. Training volume and body composition as risk factors for developing jumper's knee among young elite volleyball players. *Scandinavian journal of medicine & science in sports* 2013;23(5):607-13.
48. Team RC. 2021.

49. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. [program]. URL https://www.jstatsoft.org/v45/i03/. version, 2011.
50. Bates D, Maechler M, Bolker B, et al. Fitting Linear Mixed-Effects Models Using lme4. . *Journal of Statistical Software, 67(1), 1-48* 2015 doi: doi:10.18637/jss.v067.i01.
51. Lüdecke D. ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software* 2018;3(26):772.
52. clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. [program]. 0.5.3 version, 2021.
53. merTools: Tools for Analyzing Mixed Effect Regression Models. R package version 0.5.0. [program]. https://CRAN.R-project.org/package=merTools version, 2019.
54. rms: Regression Modeling Strategies. R package version 5.1-4. [program]. https://CRAN.R-project.org/package=rms version, 2019.
55. Bache-Mathiesen L. Training load and injury non-linearity R code GitHub: GitHub; 2021 [Available from: https://github.com/lenakba/load-injury-non-linearity-study accessed 2021-04-12.

Paper III

BMJ Open Sport & Exercise Medicine

Assessing the cumulative effect of long-term training load on the risk of injury in team sports

Lena Kristin Bache-Mathiesen ,[1] Thor Einar Andersen,[2]
Torstein Dalen-Lorentsen ,[2,3] Benjamin Clarsen,[2,4] Morten Wang Fagerland[2,5]

To cite: Bache-Mathiesen LK, Andersen TE, Dalen-Lorentsen T, *et al*. Assessing the cumulative effect of long-term training load on the risk of injury in team sports. *BMJ Open Sport & Exercise Medicine* 2022;**8**:e001342. doi:10.1136/bmjsem-2022-001342

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjsem-2022-001342).

Accepted 16 May 2022

[1]Department of Sports Medicine, Norwegian School of Sports Sciences, Oslo, Norway
[2]Department of Sports Medicine, Oslo Sports Trauma Research Centre, Norwegian School of Sports Sciences, Oslo, Norway
[3]Department of Smart Sensors and Microsystems, SINTEF Digital, Oslo, Norway
[4]Centre for Disease Burden, Norwegian Institute of Public Health, Bergen, Norway
[5]Research Support Services, Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

Correspondence to
Ms Lena Kristin Bache-Mathiesen;
l.k.bache-mathiesen@nih.no

ABSTRACT

Objectives Determine how to assess the cumulative effect of training load on the risk of injury or health problems in team sports.

Methods First, we performed a simulation based on a Norwegian Premier League male football dataset (n players=36). Training load was sampled from daily session rating of perceived exertion (sRPE). Different scenarios of the effect of sRPE on injury risk and the effect of relative sRPE on injury risk were simulated. These scenarios assumed that the probability of injury was the result of training load exposures over the previous 4 weeks. We compared seven different methods of modelling training load in their ability to model the simulated relationship. We then used the most accurate method, the distributed lag non-linear model (DLNM), to analyse data from Norwegian youth elite handball players (no. of players=205, no. of health problems=471) to illustrate how assessing the cumulative effect of training load can be done in practice.

Results DLNM was the only method that accurately modelled the simulated relationships between training load and injury risk. In the handball example, DLNM could show the cumulative effect of training load and how much training load affected health problem risk depending on the distance in time since the training load exposure.

Conclusion DLNM can be used to assess the cumulative effect of training load on injury risk.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Training load seems to affect the risk of injury in team sports.
⇒ Time since exposure to training load may determine the strength and the direction of training load's effect on injury risk.
⇒ The ability of current methodology to assess above-mentioned effects is limited.

WHAT THIS STUDY ADDS

⇒ Distributed lag non-linear models (DLNMs) were superior to all methods compared and could determine the cumulative effect of past training load.
⇒ The exponentially weighted moving average (EWMA) performed better than the rolling average and robust exponential decreasing index.
⇒ The difference between the acute:chronic workload ratio and week-to-week percentage change was negligible.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE AND/OR POLICY

⇒ Researchers can estimate the effects of training load on the risk of injury in team sports using DLNM.
⇒ More consistent methodology in training load and injury risk studies will improve comparability and reproducibility.

INTRODUCTION

In recent years, researchers have attempted to determine the effect of training load on the risk of sports injuries and other sports-related health problems.[1] Training load is the physical exertion that the athlete has been exposed to and is a combination of the exposure itself (external load) and the physiological and psychological stressors applied to the athlete in response to the exposure (internal load).[2] Relationships between risk factors and sports injuries are often complex,[3] as the effect of risk factors may depend on the presence or absence of other risk factors,[3] the current state of the athlete,[4] and they may also act non-linearly on the risk of injury.[4]

Assessing training load poses additional challenges.[5 6] It is a multidimensional construct that can be measured in multiple ways.[7] Hypotheses suggest that not only the amount of training load, but also the relative change in training load affect injury risk.[5] Balanced training load exposure may both cause and protect against injury through building fitness and fatigue.[8] A central concern is that training load is a time-varying exposure with special properties.[5 9] The training load exposure on the current day affects injury risk directly—an athlete cannot sustain a sports injury without participating in a sporting activity.[5] Training load may, however, also be a so-called time-lagged effect.[10] The training load on the previous day may contribute to the injury risk on the current day. To further add complexity, training load is likely to have a *protracted* time-lagged effect.[11] The injury risk at any given time is the result of multiple training load exposure events of different intensities sustained in the past.[12] In summary, no single training load exposure event is thought to affect injury risk in isolation, rather it is the long-term exposure to training load leading up to the event collectively that is assumed to influence injury occurrence.

To meet these assumptions, previous research has addressed some of the complexities of modelling training load statistically.[9 13] A statistical model is a generalisation that is unlikely to tailor the prognostic course of an individual accurately,[14] but it may inform researchers and clinicians about causation and patterns of injury risk. Among others, statistical solutions have been proposed to handle the time-varying effects,[9] the potential for non-linear effects,[15] the cumulative effect,[13 16] and the effect of relative training load[17] in the risk of injury. While statistical models and approaches have been recommended to handle these challenges in isolation, it is still unknown how to explore all the raised challenges in symphony. That is, accounting for time-varying effects, non-linear effects and cumulative effects simultaneously.

We aimed to determine how to model training load when assessing its cumulative effect on the risk of injury or health problems in a longitudinal team sports study.

MATERIALS AND METHODS

First, we ran a simulation study based on football data with internal training load measures to compare the performance of different statistical approaches. Then, we implemented the best performing approach on a handball dataset with training load and injury measures to demonstrate how it can be used in practice.

Football data simulation

To compare the performance of different statistical approaches, it is common to run stochastic simulations.[18] 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 (min: 16, max: 34)).[19] We used seven methods to model the relationship between training load and injury risk. To compare the performance of the seven methods, we calculated the deviation between the relationship estimated by each method and the 'true' simulated relationship (box 1, online supplemental file 1, online supplemental figure S1). More details about all methods are available in online supplemental file 2.

> **Box 1 Summary of the football data simulation**
>
> 1. Sample session rating of perceived exertion values from observed training load data in football.
> 2. Simulate time-to-event relationships between training load and injury with seven different scenarios of time-dependent effects.
> 3. Use four different methods on the absolute training load and three different methods on the relative training load to model the relationship between training load and simulated injuries.
> 4. Calculate performance measures.

Analyses and simulations were performed using R 4.1.2.[20–22] Code and data are available online.[23]

Step 1: preparing data

Internal training load was measured with the daily session rating of perceived exertion (sRPE)[24]: the duration of the activity in minutes multiplied by the player's reported perceived intensity of the activity on a scale from 0 to 10. We simulated a training load study by sampling sRPE values from the observed football dataset. The relative training load from 1 day to the next was calculated with the symmetrized percentage change (%ΔsRPE).[25] A larger study was simulated: 250 participants (10 football teams), followed for one full season (300 days).

Step 2: simulating time-to-event data

We simulated injuries under different relationship scenarios with the sampled training load. The risk of injury at any given time was predetermined with a time-to-event Cox regression model. Only one injury was simulated per individual. We use the term injury to describe the simulated events. However, the events can also be considered occurrences of pain or other health problems.

The relationship between absolute training load and injury risk was simulated to be J-shaped (online supplemental file 1 figure S2A).[15] Under this assumption, the lowest point of risk was intermediate levels of training load. The highest point of risk was set at high levels of training load.

For relative training load, we simulated a linear relationship with injury risk (online supplemental figure S2C). Higher loads on the current day compared with load on the previous day increased risk, while lower loads on the current day compared with the previous day reduced risk.[8]

In addition, we simulated the following time-dependent scenarios for both the absolute training load and the relative training load (online supplemental file S3):

- Constant. Across 4 weeks (28 days), the effect of training load has a constant effect each day.
- Decay. Across 4 weeks (28 days), the effect of training load gradually decays for each day.[13] This was hypothesised as a likely scenario if past training load has a direct effect on injury risk.
- Exponential decay. On the current day (day 0), training load has the highest risk of injury. The effect of training load drops exponentially the past 4 weeks (28 days). This was hypothesised as a likely scenario if past training load has an indirect effect on injury risk.
- Direct, then inverse. Training load values on the current week (acute) increases risk of injury, while the training load values 3 weeks before the current week (chronic) decreases risk of injury (results in supplementary).[17] This scenario represents a hypothesis that

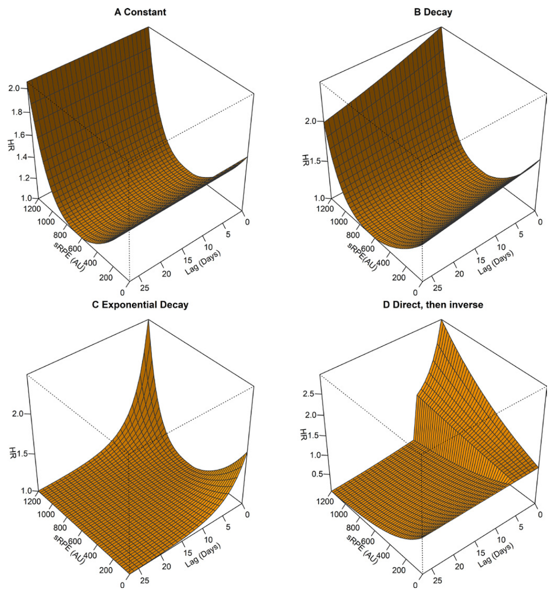

Figure 1 The four simulated relationships between absolute training load and injury risk. The relationships are a combination of the J-shaped function on the absolute training load exposure (figure 2A) and the different functions on the time since training load was sustained (online supplemental figure 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 0–6), but linearly decreases injury risk thereafter. This was the shape simulated with a linear model on the absolute training load (online supplemental figure S2B). Training load had no effect after the 27th lag day (4 weeks) in all four scenarios (not shown).

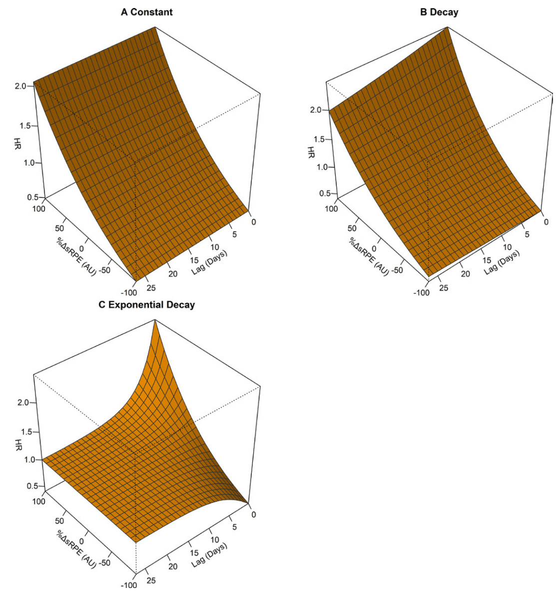

Figure 2 The three simulated relationships between relative training load and injury risk. The relationships are a combination of the linear function on the relative training load exposure (figure 2C) and the different functions on the time since training load was sustained (online supplemental figure S3). Relative training load is measured with the symmetrised percentage change (%Δ) in 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 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).

chronic load is a measure of fitness and absolute acute load is a measure of fatigue.[17] 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. Therefore, for this time-lag scenario, we simulated a linear relationship with the absolute training load, and the relative training load was not considered (online supplemental figure S2B).

In summary, seven different relationships between training load and injury risk were simulated (figures 1–2).

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. We chose the most frequently used methods in training load and injury research,[26 27] methods proposed as potential alternatives[13 16] and a method developed to handle similar challenges in epidemiology.[10] Cox regression was used to estimate the relative risk of injury, where internal training load, sRPE or %ΔsRPE was modified or modelled with different methods.

For absolute training load, we modelled the following methods with a quadratic term:

- Rolling average (RA).[28]
- Exponentially weighted moving average (EWMA).[13]
- Robust exponential decreasing index (REDI).[16]
- Distributed lag non-linear model (DLNM).[10 12]

For relative load, we modelled the following methods with a linear term:

- Week-to-week percentage change.[29]
- Acute:chronic workload ratio (ACWR),[17] 7:28 coupled RA.[30]
- DLNM.

Step 4: calculating performance measures to compare methods

We visualised the predicted cumulative risk versus the true cumulative risk in line graphs. The root-mean-squared-error (RMSE), a combined measure of accuracy and precision, was calculated between the predicted and true cumulative hazard. The lower the RMSE, the better the method. We also calculated RMSE on the predicted injury value versus the observed value (the model residuals).

The Akaike's Information Criterion (AIC) for model fit, coverage of 95% CI, average width of CI and the percentage of simulations where the methods had the lowest RMSE and lowest AIC were also calculated.

Implementation in a handball dataset

The model that performed best in our preliminary analyses of simulated data, the DLNM, was implemented on an actual data set from another team sport, to illustrate how it can be used in practice. To explore the potential for a time-dependent, cumulative effect of training load on health problem risk, we chose a Norwegian elite youth handball cohort (n=205, 36% male, mean age: 17 years (SD: 1), followed 237 days). Although the high amount of missing data (64% of sRPE values) renders it unsuitable for a study of causal inference, it had a sufficient number of health problems for the current methodology study (n=471 health problems).

RPE and duration were collected from the players after each training and match, from which daily sRPE was determined.[31] The handball players reported daily whether they had 'no health problem' or 'a new health problem'. Any response of 'a new health problem' was considered an 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.[32]

Missing sRPE data were imputed with multiple imputation.[33] Cox regression was run with health problem (yes/no) as the outcome and the DLNM of sRPE as the exposure.[9] We adjusted for sex and age as potential confounders and included a frailty term to account for recurrent events.[34] DLNM combines a function on the magnitude of sRPE and a function of the distance since day 0 up to lag 27 (4 weeks). The sRPE was modelled with restricted cubic splines[15] and the lag function with a linear model. The model predictions were visualised to assess the ability of DLNM to explore effects.

RESULTS

Football data simulation

Absolute training load

TheDLNM was the only method that discovered the simulated J-shaped relationship between absolute training load and cumulative risk of injury under all the main time-dependent effects (figure 3). It had, by far, the lowest mean external RMSE (online supplemental file 1 figure S4A-C), the lowest internal RMSE (table 1) and the lowest AIC (online supplemental figure S4D-F). Despite consistently having the narrowest average CI width (≈2 vs >3 (all other methods)), it also had the second-to-highest coverage of 95% CIs under the constant scenario and the highest under the decay scenario (table 1). Except for the exponential decay scenario, all methods had poor coverage overall (<=35%, table 1).

The EWMA was able to detect the exponential decay scenario (figure 3J) and had better accuracy than the rolling average and the robust exponential decreasing index for the decay scenario (figure 3E–G). It had the lowest mean external RMSE and AIC of all three scenarios and methods (table 1, online supplemental figure S4), although, under the constant scenario, the CIs reached negative values (figure 3B).

The rolling average was able to model the constant scenario (figure 3A) and had a mean internal RMSE of 0.113547, slightly lower than EWMA at 0.113548. Under this condition, it had the second best (rank 2) external RMSE in 31% of simulations and third best (rank 3) in 52% of simulations, with similar results for AIC (31% rank 2, 58% rank 3; online supplemental table S1). Here, EWMA was most frequently ranked second best for RMSE and AIC (45% and 39%, respectively (online supplemental table S1).

REDI had consistently the highest mean external RMSE and AIC (online supplemental figure S4, table 1). It was most frequently rank 4 for external RMSE under the constant and decay scenarios and for AIC under all scenarios (online supplemental table S1). Furthermore, REDI consistently had the lowest coverage of 95% CIs (table 1). Instead of discovering that high levels of absolute training load increases injury risk, REDI estimated that high absolute training load decreases injury risk under the exponential decay scenario (figure 3K).

No method was able to accurately model the direct, then inverse scenario (coverage=0%, online supplemental figure S5, online supplemental table S2).

Relative training load

The Distributed Lag Non-Linear Model (DLNM) was also capable of discovering the cumulative hazard of injury for relative training load (figure 4C, F, I). It had the lowest mean internal RMSE and AIC for the Constant and Decay

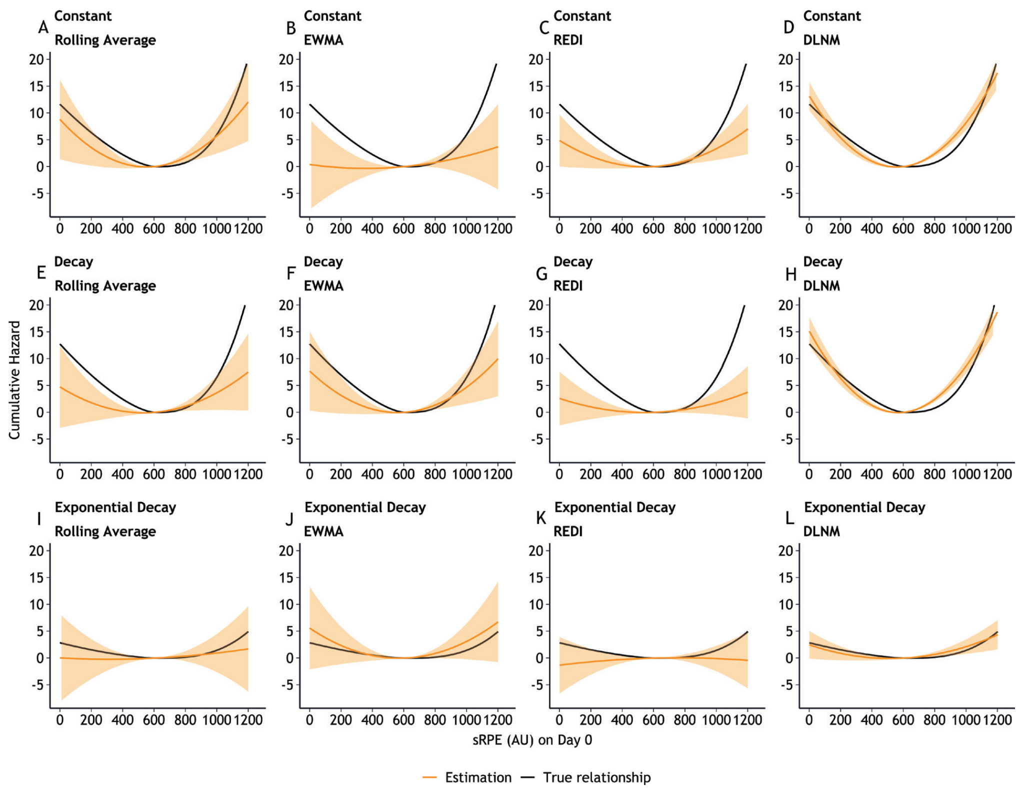

Figure 3 The relationship between absolute training load measured by the session rating of perceived exertion (sRPE) in arbitrary units (AUs) and the risk of injury on the current day (day 0) estimated by four different methods (yellow line), compared with the simulated, true relationship (black line). The y-axis denotes the cumulative hazard – the sum of all instantaneous risks of injury from the past up until the current day. Relationships were simulated under different scenarios, (A–D) constant: the risk of absolute training load is constant over time; (E–H) decay: the effect of absolute training load was at its highest on the current day (day 0) and reduced linearly for each lag day back in time; (I–L) exponential decay: the risk of absolute training load was at its highest on the current day (day 0) and reduced exponentially for each lag day back in time. Methods used to detect these effects were the rolling average, the exponential weighted moving average (EWMA), the robust exponential decreasing index (REDI), and the distributed lag non-linear model (DLNM). Yellow bands are 95% CIs. The figure shows one random simulation of 1900 performed.

scenarios (online supplemental figure S6), but for the Exponential Decay scenario, it had the lowest mean AIC and highest internal RMSE (table 1, online supplemental figure S6). Under all scenarios, DLNM had the lowest AIC in nearly 100% of simulations (online supplemental table S3). Although it was most frequently rank 1 internal RMSE for the Constant (52% of simulations) and Decay scenarios (57% of simulations), the rankings varied, and the Acute:Chronic Workload Ratio and Week-to-week %Δ were rank 1 ~23% of the time each (online supplemental table S3).

The Acute:Chronic Workload Ratio (ACWR) and week-to-week %Δ failed to discover a relationship between training load and injury under the Constant scenario (figure 4A, B). ACWR did not find a relationship under the Exponential Decay scenario, either (figure 4G). Both methods had wide confidence intervals, and ACWR fanned to higher uncertainty under higher levels of acute training load relative to chronic training load (figure 4). ACWR had marginally lower internal RMSE and lower AIC than week-to-week %Δ (table 1), and was rank 2 slightly more frequently than rank 3 (online supplemental table S3), except under the Exponential Decay scenario where the opposite was the case.

Handball example data analysis

The Distributed Lag Non-linear Model indicated, with high uncertainty, an increased risk of a health problem on the current day (HR (HR)>=1.2) for players with high internal load (sRPE above 4 000, figure 5A). This tapered to no effect if the training load was performed around a week ago (6 days before the current day, figure 5D), to a decreased risk of health problems the further in the past high training loads were sustained, to a HR of 0.75 on the

Table 1 Mean performance of methods used to estimate the effect of training load on injury risk (n simulations=1900).

Relationship	Method	External RMSE*	Internal RMSE	AIC	Coverage (%)	AW	Coverage MCSE
Absolute training load							
Constant	Rolling average	4.85	0.113547	1422.92	34.7	5.17478	0.90
	EWMA	4.77	0.113548	1423.42	36.3	5.17179	0.91
	REDI	5.53	0.113557	1424.10	20.3	3.40114	0.74
	DLNM	1.44	0.112434	1317.15	34.8	2.05600	0.95
Decay	Rolling average	5.38	0.113590	1421.80	30.2	5.16930	0.87
	EWMA	5.17	0.113587	1421.85	31.8	5.12554	0.88
	REDI	6.21	0.113605	1423.80	18.7	3.42154	0.71
	DLNM	1.55	0.112245	1295.30	32.4	2.07977	0.93
Exponential decay	Rolling average	2.13	0.113599	1424.65	85.0	5.54695	0.58
	EWMA	1.88	0.113588	1423.86	85.1	5.37141	0.61
	REDI	1.97	0.113603	1425.00	74.2	3.69208	0.64
	DLNM	0.76	0.113368	1407.08	81.6	2.02633	0.65
Relative training load (%Δ)†							
Constant	ACWR		0.113643	1426.16			
	Week-to-week %Δ		0.113646	1426.40			
	DLNM %Δ		0.113627	1389.28			
Decay	ACWR		0.113615	1424.73			
	Week-to-week %Δ		0.113617	1425.12			
	DLNM %Δ		0.113553	1383.52			
Exponential decay	ACWR		0.113565	1423.33			
	Week-to-week %Δ		0.113566	1423.27			
	DLNM %Δ		0.113700	1401.39			

*Monte Carlo SE for RMSE was <0.001 for all simulations. The scale of the RMSE depends on the scale of the coefficients, and it is therefore only interpretable by comparing values in the same analysis – the values cannot be interpreted in isolation.
†Due to differences in scale between methods and simulation for relative training load, external RMSE, coverage, and AW could not be calculated in a comparable manner.
ACWR, acute:chronic workload ratio; AIC, Akaike's information criterion; AW, average width of 95% CIs; Coverage, coverage of 95% CIs; DLNM, distributed lag non-linear mode; EWMA, exponentially weighted moving average; MCSE, Monte Carlo Standard Error; REDI, robust exponential decreasing index; RMSE, root-mean-squared error.

27th day before the current day (figure 5B). The cumulative risk was increased if an individual performed no training in the past and had high internal training load on the current day (figure 5C). None of the effects were significant (p>=0.8) and confidence intervals were broad (online supplemental table S4).

DISCUSSION

This is the first simulation study to explore methods for assessing the cumulative effect of long-term training load on injury or health problem risk in team sports. The Distributed Lag Non-linear Model (DLNM) had the highest combined accuracy and precision, the highest certainty, and the best model fit for almost all studied scenarios. It was the only method capable of exploring both the effects of the magnitude of training load and the time-dependent effects of past training load exposure.

In the application of DLNM on a handball cohort, we were hampered by poor data quality. Also, due to anonymization, few covariates were available for confounder adjustment. The effects may have been spurious. We have included the analysis only as an illustration of how to use the DLNM in practice.

Modelling methods for absolute training load

For determining the cumulative effect of the absolute training load, the Rolling Average was outclassed by the Exponentially Weighted Moving Average (EWMA). When the effect of absolute training load was simulated to be the same regardless of the distance in time since the current day – the scenario in which Rolling Average was thought to be appropriate – Rolling Average was only marginally better than the EWMA. EWMA had a better fit under the more realistic scenarios where the effects of training load decayed based on distance in time,

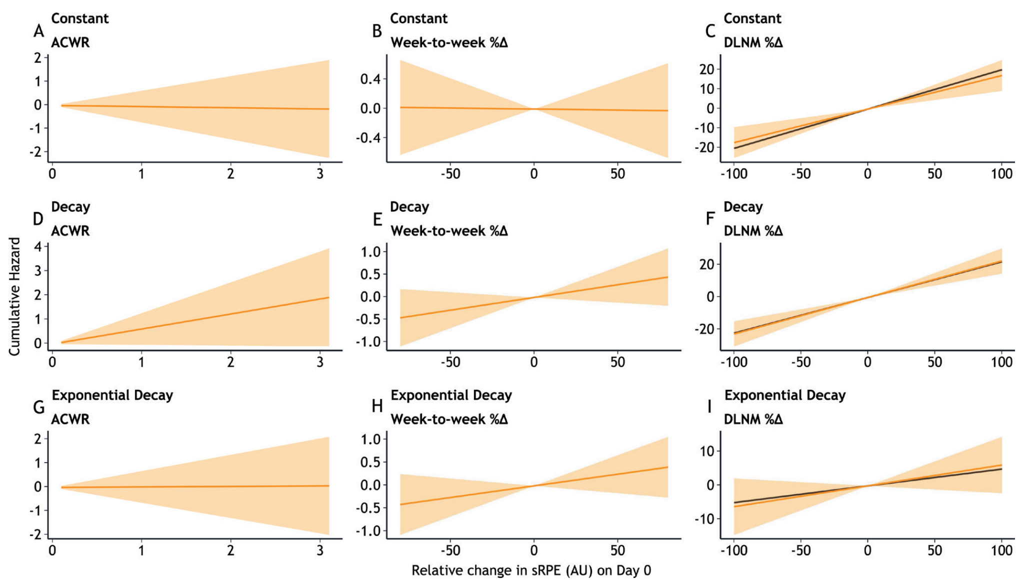

Figure 4 The relationship between relative training load measured in the daily percentage change of session rating of perceived exertion (sRPE) in arbitrary units (AUs) and the risk of injury on the current day (day 0) is estimated by three different methods (yellow line). The y-axis denotes the cumulative hazard – the sum of all instantaneous risks of injury from the past up until the current day. Relationships were simulated under different scenarios, (A–C) constant: the risk of relative training load was constant over time; (D–F) decay: the effect of relative training load was at its highest on the current day (day 0) and reduced linearly for each lag day back in time; (G–I) exponential decay: the risk of relative training load was at its highest on the current day (day 0) and reduced exponentially for each lag day back in time. Methods used to detect these effects were the acute:chronic workload ratio (ACWR), the week-to-week percentage change (%Δ) and the distributed lag non-linear model (DLNM) on daily percentage change Δ%. The DLNM, being on the same scale as the simulation, is also compared with the true, simulated relationship (black line). Yellow bands are 95% CIs. The figure shows one random simulation of 1900 performed.

both linearly and exponentially. This is in line with the concerns raised by Menaspà,[28] that the rolling average fails to take into account that training load performed in the past contributes less to injury risk than recent training load

The Robust Exponential Decreasing Index (REDI) was also outperformed by EWMA, under both scenarios where the training load effect decayed based on distance in time. Across the board, REDI had the highest RMSE, highest AIC, and lowest coverage of 95% confidence intervals. Although it had better RMSE under the Exponential Decay scenario than the rolling average, it erroneously estimated that higher internal training loads decreased injury risk (inverse relationship), when it was actually the opposite (ie, higher training load increased injury risk). REDI has previously been compared on observed training load values where the true relationship between training load and injury was unknown,[35] and it was recommended for its ability to handle missing data.[16] We believe that using imputation methods is more suitable for longitudinal data,[33] and in such cases, the advantage of specifying weights on missing observations is no longer applicable. REDI was among the methods that do not require a full time period (ie, 28 days) before the first calculation, but for comparability, we had to run it with the same limitation as the other methods. Arguably, it may therefore have performed better in a real study. On the other hand, this would also have been the case for the Distributed Lag Non-Linear Model (DLNM), which was vastly superior to all other methods analysed, even with this constraint.

DLNM had the lowest mean RMSE, AIC, and narrowest 95% CI width compared with the other three methods for all scenarios. The DLNM was the only method that did not require subjective aggregation. Aggregation distillates the information available in the data to a summary, and these summaries are all the Cox regression model must work with. This increases the uncertainty of the estimates. In contrast, DLNM uses all the information available in the data.[12] Furthermore, no subjective determination of time-lag weights is required. Using splines or fractional polynomials, it can explore non-linearity in both the magnitude of the effect of absolute training load and in the time-dependent effects.[15]

While it performed best compared with other methods, DLNM was unable to model the "Direct, then inverse" scenario. This scenario was built on the theory that training load exposure the current week increase risk

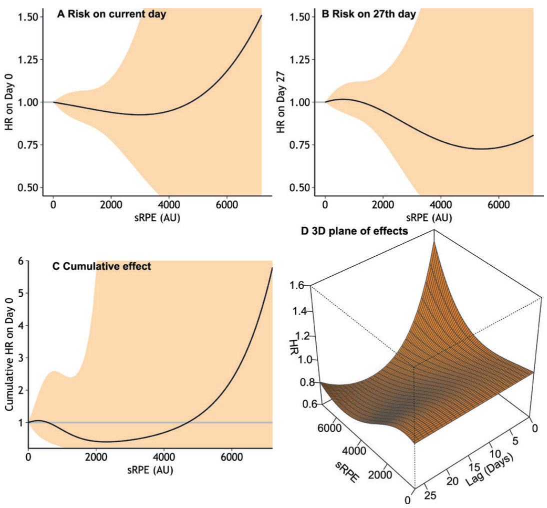

Figure 5 Explorations of the relationship between training load and the risk of suffering a health problem in a Norwegian elite youth handball cohort. Training load is measured by the session rating of perceived exertion (sRPE) in arbitrary units (AUs), shown on all x-axes. The health problem risk is measured by the Hazard Ratio (HR). HR >1 indicates an increased instantaneous health problem risk compared with an individual who had no training load (sRPE=0), <1 a decreased risk. Figure part A shows the risk of a health problem on the y-axis for each level of sRPE on the x-axis, given that the sRPE is sustained on the current day (day 0). Figure part B shows the same figure, given that the sRPE is sustained on the 27th lag day (4 weeks prior). Figure part C shows the cumulative HR – the collective risk of a health problem on the current day given the sRPE sustained in all the days prior to the current day. Finally, figure part D shows the risk relationship between absolute training load (sRPE) on the x-axis and the time since the training was sustained (lag) on the y-axis, where 0 is the current day and 27 is 4 weeks in the past. Risk in HR is on the z-axis. Yellow bands in (A–C) are the 95% CIs surrounding the estimates. The predictions pertain to a 17-year-old female. Based on 471 health problems from 205 handball players.

while those sustained the previous 3 weeks reduce risk.[8] Higher sample sizes than those in the current simulation may be needed to discover such a complex shape, if it were to exist. The splines may have required more than three knots, and linear splines may have been a better option than cubic splines to discover the sudden change in direction of effect.

Modelling methods for relative training load

Studying the relative training load proved challenging, as all methods compared were on different scales. According to the AIC, the most comparable metric,[12] DLNM had the best model fit under all scenarios. Given that we simulated an effect on the risk of injury based on the symmetrized percentage change from 1 day to the next, this was to be expected. The week-to-week percentage change and ACWR assume that day-to-day differences are of little to no importance. Currently, the time-period of relative training load that is relevant towards injury risk is debatable[36]; a calendar week may be arbitrary for many sports. We argue that if DLNM can detect the effect of day-to-day relative change, it should be flexible enough to detect less granular effects. In particular, team sports such as football often operate in micro-cycles of days since the previous match up to and including the next match.[15] However, it would still be up to the researcher to calculate percentage changes on time periods of their choosing before running DLNM, with the inherent difficulties of ratios.[25]

Even with the symmetrized percentage change, the percentage change cannot be calculated if the numerator or denominator is zero. Recovery days are an important aspect of training load history and must be evaluated to fully understand the effects of training load. This is a challenge that remains unsolved.

An application of distributed lag non-linear models in handball

The Distributed Lag Non-linear Model was able to explore non-linear time-dependent effects in the observed Norwegian youth elite handball data. The results had a high degree of uncertainty (p>=0.8), and we caution against considering them as evidence of a causal or associative relationship. They nevertheless illustrate how DLNM can be used in practice. DLNM can show how different levels of training load affects risk, and also how the effects changes with the distance in time since the training load exposure. It can also show the combined effect of these two dimensions and estimate the cumulative effect. However, performing DLNM and the corresponding visualisations in a training load and injury or health problem risk study may require collaboration with a statistician.[37] In addition, large sample sizes and good data quality may be needed to meet the complexity of the training load and injury risk relationship. In the handball data, 471 health problems occurred in 205 participants. As this was insufficient, future research may require even more participants for an accurate measure of effect.

Limitations

To feasibly analyse all results in a single article, we had to limit the number of methods compared in the simulations. This meant that two recently-proposed methods of relative training load were not among the compared methods.[38 39] Additionally, different variants of the ACWR were not considered, as these have been explored extensively in other studies.[30 40]

All methods in the simulation were run with the same specification for all scenarios to ensure consistency and comparability. In a real study, clinical rationale and hypothesis, as well as sensitivity analyses of model fit, would aid in determining the number and location of knots in splines for DLNM, the lambda value for EWMA and REDI, and the time-periods for RA, EWMA, REDI and ACWR. Therefore, the flexibility of methods was not fully explored. In addition, for the relative training load, the simulation assumed that daily differences had an effect, an assumption that favoured DLNM, which has superior flexibility compared with the other methods. This advantage may be less prominent if stricter assumptions (ie, differences at the micro-cycle level) can be made[15]; however, we believe that the flexibility of the DLNM is one of its greatest strengths, rendering it useful in a wide range of situations.

CONCLUSION

The Distributed Lag Non-Linear Model is ideal for exploring the cumulative effect of the absolute training load and relative training load on injury risk, while accounting for time-dependent effects. For causal studies where training load is not the exposure of interest, but a confounder in need of adjustment, using the Exponentially Weighted Moving Average for the absolute training load is an alternative.

Twitter Lena Kristin Bache-Mathiesen @lena_kbm and Torstein Dalen-Lorentsen @torsteindalen

Acknowledgements We thank Christian Thue Bjørndal for granting access to the Norwegian elite youth handball data. We also thank Garth Theron for programming the Norwegian Premier League database. This research would not be made possible without the collaboration of coaches and athletes, and we thank the participants who contributed.

Contributors All authors conceptualized the study. Authors TEA, TD-L, and BC determined the aim and scope of the study, and the study design, with input from LKB-M and MWF. Author LKB-M performed simulations, statistical analyses, and wrote the manuscript under supervision from authors MWF and TEA. All authors contributed with notable critical appraisal of the text and approved the final version. LKB-M was the guarantor of this study.

Funding The Oslo Sports Trauma Research Centre provided all funding for performing this study. The Oslo Sports Trauma Research Center has been established at the Norwegian School of Sport Sciences through generous grants from the Royal Norwegian Ministry of Culture, the South-Eastern Norway Regional Health Authority, the International Olympic Committee, the Norwegian Olympic Committee & Confederation of Sport, and Norsk Tipping AS.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by The Norwegian Premier League football study and Norwegian elite youth handball study was approved by the Ethical Review Board of the Norwegian School of Sport Sciences, and by the Norwegian Centre for Research Data, with registration numbers 722773, and 407930, respectively. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information. All data relevant to the study are included in the article, are available as supplementary files, or available upon reasonable request. The anonymous training load variable from the Norwegian Premier League football data, and all statistical programming code, is available in a GitHub repository. The anonymised Norwegian elite youth handball data are available upon reasonable request.

ORCID iDs
Lena Kristin Bache-Mathiesen http://orcid.org/0000-0002-9466-7003
Torstein Dalen-Lorentsen http://orcid.org/0000-0003-4062-7601

REFERENCES

1 Griffin A, Kenny IC, Comyns TM. The Association Between the Acute:Chronic Workload Ratio and Injury and its Application in Team Sports: A Systematic Review. *Sports Med* 2019:1–20.
2 Soligard T, Schwellnus M, Alonso J-M, *et al*. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br J Sports Med* 2016;50:1030–41.
3 Bittencourt NFN, Meeuwisse WH, Mendonça LD, *et al*. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition-narrative review and new concept. *Br J Sports Med* 2016;50:1309–14.
4 Stern BD, Hegedus EJ, Lai Y-C. State dependence: does a prior injury predict a future injury? *Phys Ther Sport* 2021;49:8–14.
5 Windt J, Gabbett TJ. How do training and competition workloads relate to injury? the workload-injury aetiology model. *Br J Sports Med* 2017;51:428–35.
6 Kalkhoven JT, Watsford ML, Coutts AJ, *et al*. Training load and injury: causal pathways and future directions. *Sports Med* 2021;51:1137–50.
7 Impellizzeri FM, Marcora SM, Coutts AJ. Internal and external training load: 15 years on. *Int J Sports Physiol Perform* 2019;14:270–3.
8 Gabbett TJ. The training-injury prevention paradox: should athletes be training smarter and harder? *Br J Sports Med* 2016;50:273–80.
9 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;53:61–8.
10 Bhaskaran K, Gasparrini A, Hajat S, *et al*. Time series regression studies in environmental epidemiology. *Int J Epidemiol* 2013;42:1187–95.
11 Richardson DB. Latency models for analyses of protracted exposures. *Epidemiology* 2009;20:395–9.
12 Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear models. *Stat Med* 2014;33:881–99.
13 Williams S, West S, Cross MJ, *et al*. Better way to determine the acute:chronic workload ratio? *Br J Sports Med* 2017;51:209–10.
14 Nielsen RO, Bertelsen ML, Møller M, *et al*. Methods matter: exploring the 'too much, too soon' theory, part 1: causal questions in sports injury research. *Br J Sports Med* 2020;54:1119–22.
15 Bache-Mathiesen LK, Andersen TE, Dalen-Lorentsen T, *et al*. Not straightforward: modelling Non-linearity in training load and injury research. *BMJ Open Sport Exerc Med* 2021;7:e001119.
16 Moussa I, Leroy A, Sauliere G, *et al*. Robust exponential decreasing index (REDI): adaptive and robust method for computing cumulated workload. *BMJ Open Sport Exerc Med* 2019;5:e000573.
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. *Br J Sports Med* 2016;50:471–5.
18 Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med* 2019;38:2074–102.
19 Theron GF. *The use of data mining for predicting injuries in professional football players*. University of Oslo, 2020.
20 Team RC. *R: a language and environment for statistical computing*. Vienna, Austria: Foundation for Statistical Computing, 2021.
21 Sylvestre M-P, Abrahamowicz M. Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Stat Med* 2009;28:3437–53.

22 Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. *J Stat Softw* 2011;43:1.
23 Bache-Mathiesen LK. Modelling the cumulative effect of training load on injury risk: R code Repository, 2022. Available: https://github.com/lenakba/modelling-training-load
24 Foster C, Florhaug JA, Franklin J, *et al*. A new approach to monitoring exercise training. *J Strength Cond Res* 2001;15:115.
25 Curran-Everett D, Williams CL. Explorations in statistics: the analysis of change. *Adv Physiol Educ* 2015;39:49–54.
26 Dalen-Lorentsen T, Andersen TE, Bjørneboe J, *et al*. A cherry, ripe for Picking: the relationship between the Acute-Chronic workload ratio and health problems. *J Orthop Sports Phys Ther* 2021;51:162–73.
27 Wang A, Healy J, Hyett N, *et al*. A systematic review on methodological variation in acute:chronic workload research in elite male football players. *Sci Med Footb* 2021;5:18–34.
28 Menaspà P. Are rolling averages a good way to assess training load for injury prevention? *Br J Sports Med* 2017;51:618.1–9.
29 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. *Int J Sports Sci Coach* 2021;16:370–9.
30 Coyne JOC, Nimphius S, Newton RU, *et al*. Does mathematical coupling matter to the acute to chronic workload ratio? A case study from elite sport. *Inter J f Spo Physi Perfor* 2019;14:1447–54.
31 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. *Front Sports Act Living* 2021;3:635103.
32 Fuller CW, Ekstrand J, Junge A, *et al*. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Scand J Med Sci Sports* 2006;16:83–92.
33 Bache-Mathiesen LK, Andersen TE, Clarsen B, *et al*. Handling and reporting missing data in training load and injury risk research. *Sci Med Foot* 2021;55:1–13.
34 Ullah S, Gabbett TJ, Finch CF. Statistical modelling for recurrent events: an application to sports injuries. *Br J Sports Med* 2014;48:1287–93.
35 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. *Front Physiol* 2020;11:1034.
36 West S, Williams S, Cazzola D. 119 The acute:chronic workload ratio: why one size does not fit all. *Br J Spor Med* 2021;55:A48.
37 Sainani KL, Borg DN, Caldwell AR, *et al*. Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. *Br J Sports Med* 2021;55:118–22.
38 Afonso J, Nakamura FY, Canário-Lemos R, *et al*. A novel approach to training Monotony and Acute-Chronic workload index: a comparative study in soccer. *Front Sports Act Liv* 2021;3:149.
39 Tysoe A, Moore IS, Ranson C, *et al*. Bowling loads and injury risk in male first class county cricket: Is 'differential load' an alternative to the acute-to-chronic workload ratio? *J Sci Med Sport* 2020;23:569–73.
40 Wang C, Stokes T, Vargas JT, *et al*. Predicting Injury Risk Over Changes in Physical Activity in Children Using the Acute:Chronic Workload Ratio. *Am J Epidemiol* 2022;191:665–73.

Supplementary I: Results

FIGURES

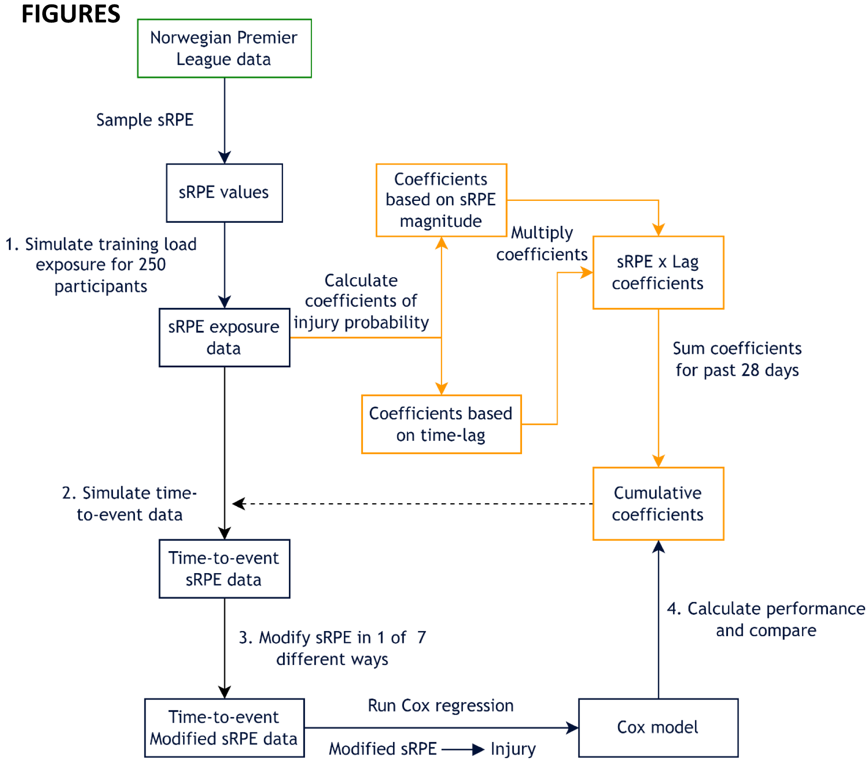

Figure S1. Summary of the simulation workflow. In Step 1, training load exposure measured by session Rating of Perceived Exertion (sRPE) was extracted from the Norwegian Premier League dataset and used to simulate training load exposure for 250 participants across 300 days. In Step 2, injury probabilities were calculated based on the cumulative training load observed the last 28 days; a combination of effect from both the magnitude of the training load (level of sRPE or %ΔsRPE) and the time since the training load occurred. Injuries were simulated based on these probabilities to generate time-to-event data. In Step 3, the absolute and relative training load exposures were modified and modelled in seven different Cox regression models. Finally, in Step 4, performance measures were calculated, and the accuracy of the different Cox models to detect the simulated relationship was assessed. Steps 1–4 were repeated 1 900 times for each of seven different simulated relationships (four for sRPE and three for %ΔsRPE) and each of seven methods.

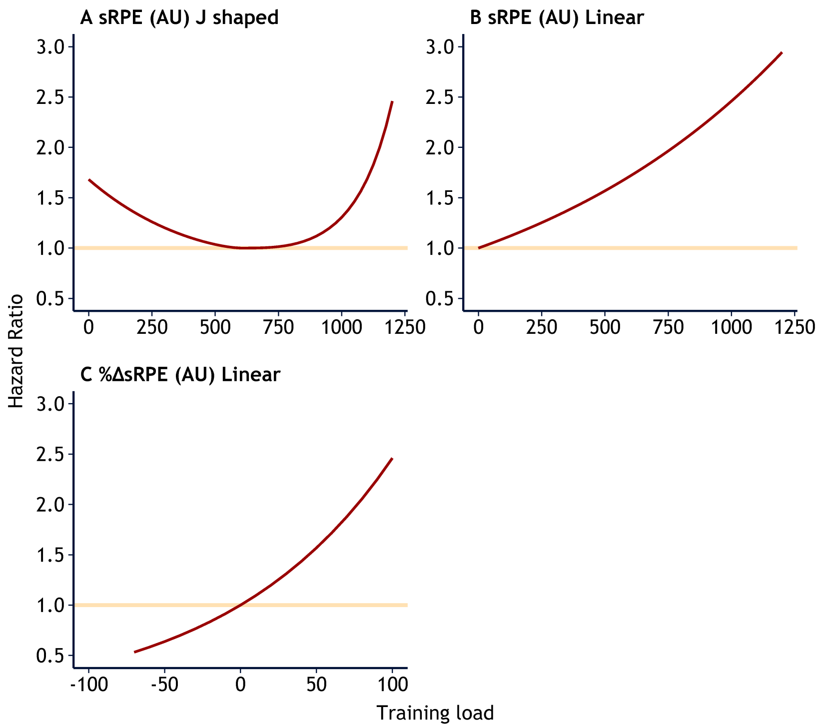

Figure S2. The simulated relationships between training load and injury risk, independent of the time since the training load exposure. Injury risk is measured by the Hazard Ratio (HR), where values > 1 (above the yellow line) indicates an increased risk and values < 1 (below the yellow line) indicates a decreased risk. Shown for (A–B) the absolute training load measured by the session Rating of Perceived Exertion (sRPE) measured in Arbitrary Units (AU), and (C) the relative training load compared to the previous day measured by the symmetrized percentage difference (%Δ) in sRPE. The absolute training load exposure was simulated with two different relationships, one J-shaped (A), and one linear (B).

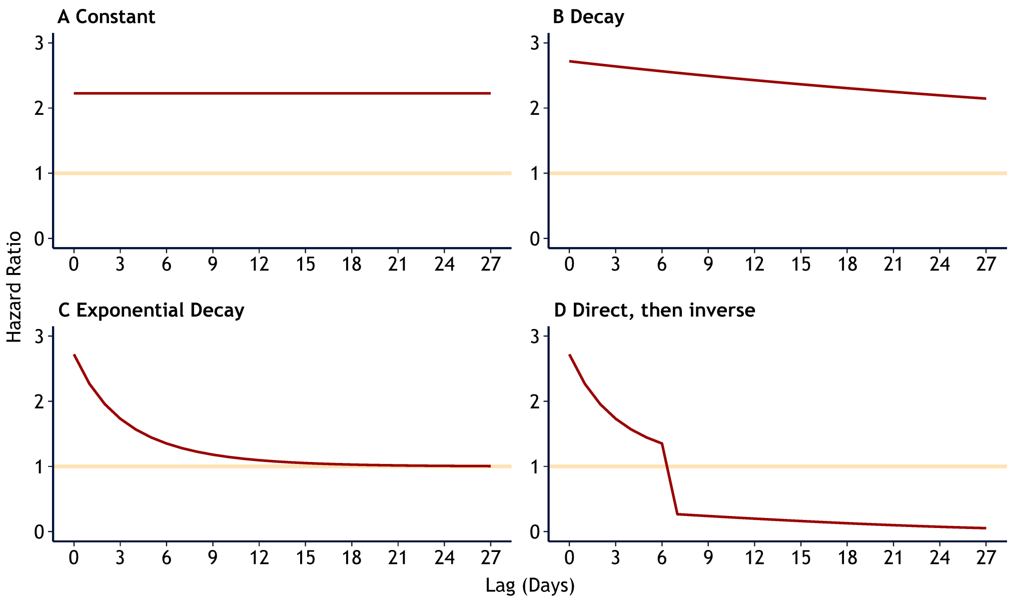

Figure S3. The simulated relationships between the time since current day (Day 0) that the training load exposure was sustained, and injury risk. Injury risk is measured by the Hazard Ratio (HR), where values > 1 (above the yellow line) indicates an increased risk and values < 1 (below the yellow line) indicates a decreased risk. The four risk shapes were (A) Constant, where the risk of training load is constant over time; (B) Decay, where the effect-size of the effect of training load is at its highest on the current day (Day 0) and is reduced for each lag day back in time; (C) Exponential Decay, where the 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 increases injury risk during the current week (Day 0–Day 6), but decreases injury risk thereafter. Training load had no effect after the 27th lag day (4 weeks) in all four scenarios (not shown).

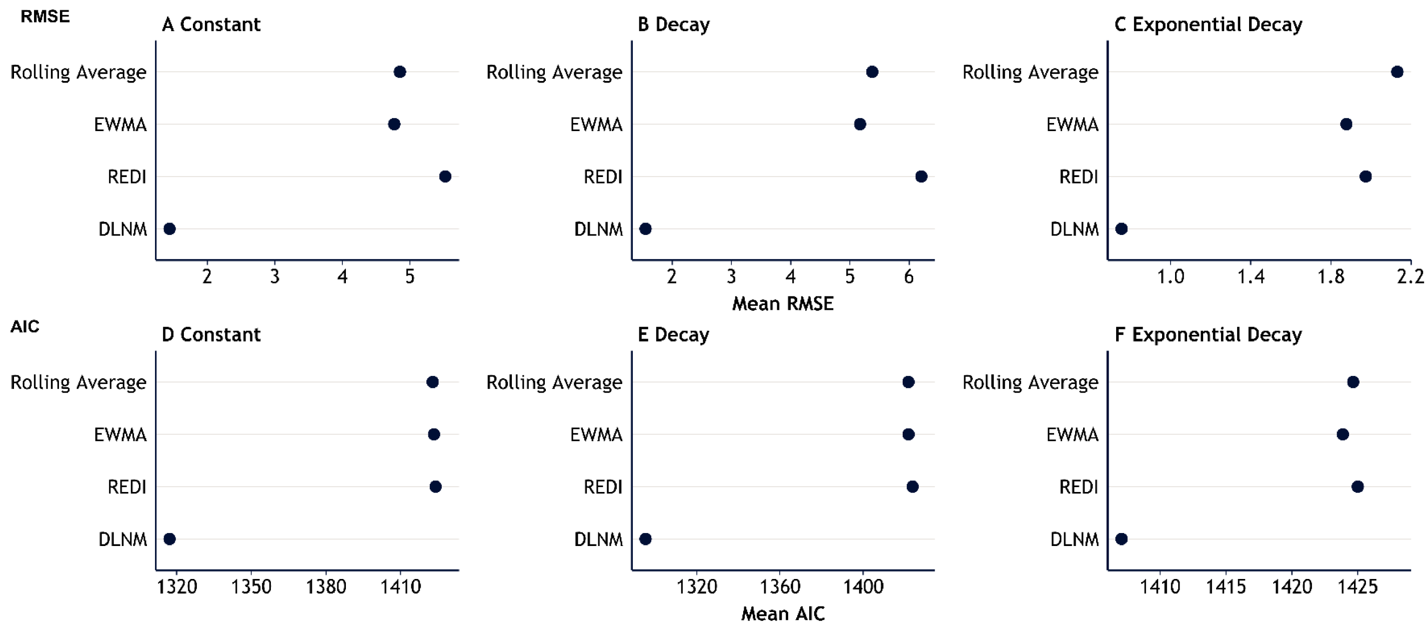

Figure S4. The mean Root-Mean-Squared Error (RMSE) and mean Akaike's Information Criterion (AIC) across 1 900 simulations of estimating the effect of absolute training load on injury risk. Due to variation in the effect sizes, calculations yield different scales for RMSE and AIC (x-axis) between relationship shapes; they cannot be compared between the three shapes, only within each shape. EWMA = Exponentially Weighted Moving Average; DLNM = Distributed Lag Non-Linear Model; REDI = Robust Exponential Decreasing Index. RMSE is calculated on the difference between the predicted risk and the simulated, true risk (External RMSE).

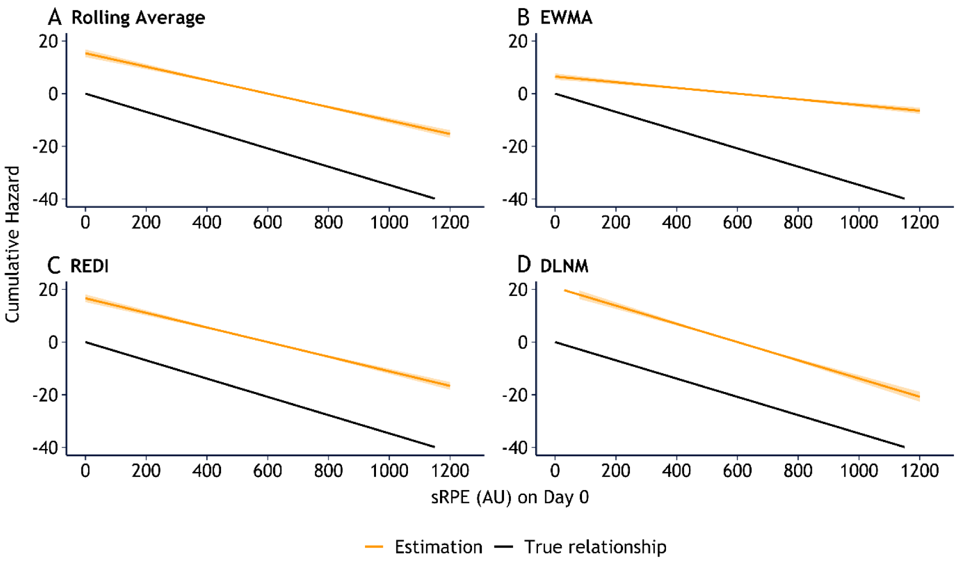

Figure S5. The relationship between absolute training load measured by the session Rating of Percieved Exertion (sRPE) in arbitrary units (AU) and the risk of injury on the current day (Day 0) estimated by four different methods (yellow line), compared with the simulated, true relationship (black line). The relationship scenario was "Direct, then inverse", where training load increases injury risk during the current week (Day 0–Day 6), but decreases injury risk thereafter (Day 7–Day 27). The Y axis denotes the cumulative hazard – the sum of all instantaneous risks of injury from the past up until the current day. Methods used to detect these effects were (A) the Rolling Average, (B) the Exponential Weighted Moving Average (EWMA), (C) The Robust Exponential Decreasing Index (REDI), and (D) the Distributed Lag Non-Linear Model (DLNM). Yellow bands are 95% confidence intervals. The figure shows 1 random simulation of 1 900 performed.

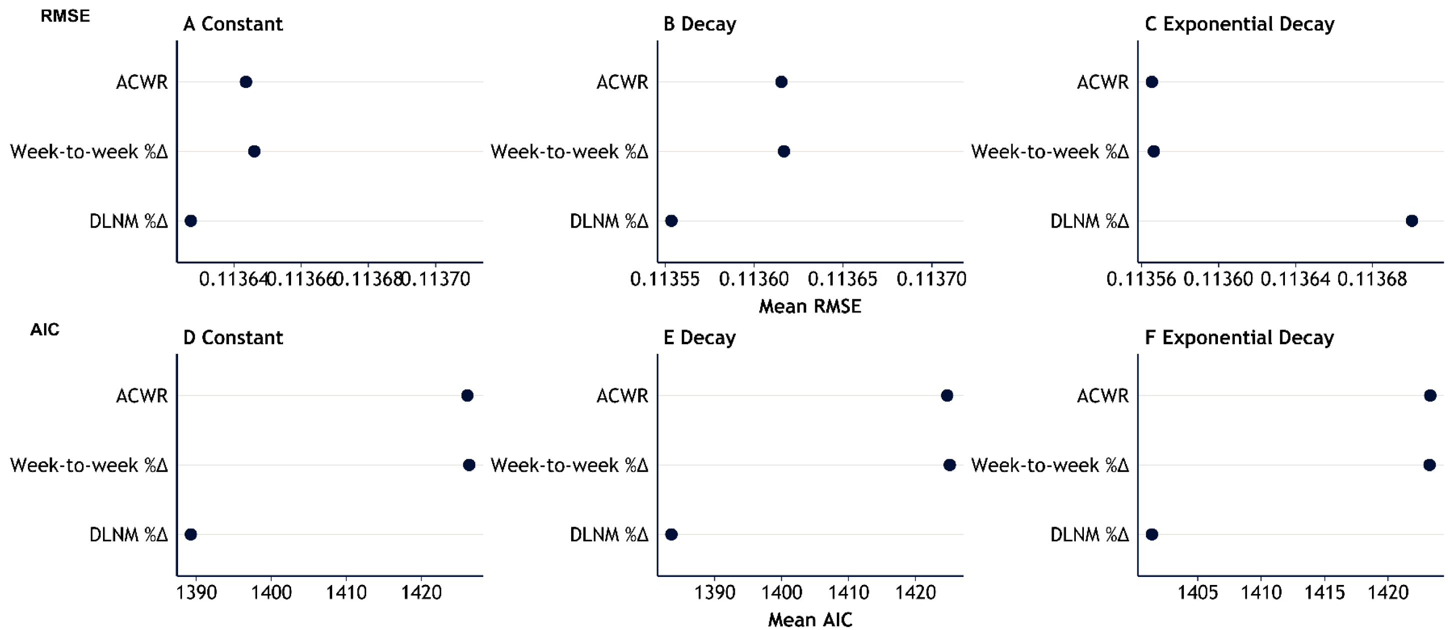

Figure S6. The mean Root-Mean-Squared Error (RMSE) and mean Akaike's Information Criterion (AIC) across 1 900 simulations of estimating the effect of relative training load on injury risk. Due to variation in the effect sizes, calculations yield different scales for RMSE and AIC (x-axis) between relationship shapes; they cannot be compared between the three shapes, only within each shape. ACWR = Acute:Chronic Workload Ratio; DLNM = Distributed Lag Non-Linear Model. RMSE is calculated on the model residuals (Internal RMSE).

TABLES

Table S1. The percentage of 1 900 simulations where methods of absolute training load had the lowest RMSE and AIC (Rank 1), had the 2nd lowest RMSE and AIC (Rank 2), and so on.

Metric	Lag scenario	Rank	Rolling Average (%)	EWMA (%)	REDI (%)	DLNM (%)
RMSE	Constant	1	2	1	0	97
		2	31	45	22	2
		3	52	27	21	1
		4	15	27	58	0
	Decay	1	1	1	0	98
		2	29	48	21	2
		3	54	26	19	0
		4	15	25	60	0
	Exponential Decay	1	11	13	13	63
		2	19	28	26	27
		3	36	27	29	8
		4	34	31	32	3
	Direct, then inverse	1	0	0	1	99
		2	0	0	99	1
		3	100	0	0	0
		4	0	100	0	0
AIC	Constant	1	0	0	0	100
		2	31	39	31	0
		3	58	24	18	0
		4	11	38	51	0
	Decay	1	0	0	0	100
		2	31	45	24	0
		3	59	24	17	0
		4	10	31	59	0
	Exponential Decay	1	1	1	1	97
		2	19	52	28	2
		3	55	22	23	0
		4	26	25	48	1
	Direct, then inverse	1	0	0	0	100
		2	0	0	100	0
		3	100	0	0	0
		4	0	100	0	0

Abbreviations: AIC = Akaike's Information Criterion; EWMA = Exponentially Weighted Moving Average; DLNM = Distributed Lag Non-Linear Model; REDI = Robust Exponential Decreasing Index; RMSE = Root-Mean-Squared Error

Table S2. Mean performance of methods used to estimate the effect of absolute training load on injury risk under the "Direct, then inverse" scenario.

	Rolling Average	EWMA	REDI	DLNM
External RMSE[1]	21.1	22.6	20.9	20.8
Internal RMSE	0.111	0.113	0.106	0.101
AIC	1116	1373	910	790
Coverage[1]	0%	0%	0%	0%
AW	1.48	1.25	1.56	1.94

Abbreviations: AIC = Akaike's Information Criterion; AW = Average Width of 95% confidence intervals; Coverage = Coverage of 95% confidence intervals; EWMA = Exponentially Weighted Moving Average; DLNM = Distributed Lag Non-Linear Model; REDI = Robust Exponential Decreasing Index; RMSE = Root-Mean-Squared Error
[1] Monte Carlo Standard Error was < 0.001 for RMSE, and 0.5 for coverage of 95% confidence intervals for all methods.

Table S3. The percentage of 1 900 simulations where methods of relative training load had the lowest RMSE and AIC (Rank 1), had the 2nd lowest RMSE and AIC (Rank 2), and so on.

Metric	**Lag scenario**	**Rank**	**ACWR (%)**	**Week-to-week %Δ (%)**	**DLNM %Δ (%)**
RMSE	Constant	1	25	23	52
		2	49	49	2
		3	26	29	46
	Decay	1	23	21	57
		2	50	48	2
		3	28	31	41
	Exponential Decay	1	31	29	41
		2	48	50	2
		3	22	21	57
AIC	Constant	1	0	0	100
		2	56	44	0
		3	44	56	0
	Decay	1	0	0	100
		2	59	41	0
		3	41	59	0
	Exponential Decay	1	1	1	99
		2	49	51	0.5
		3	52	49	0.9

Abbreviations: ACWR = Acute:Chronic Workload Ratio; AIC = Akaike's Information Criterion; DLNM = Distributed Lag Non-Linear Model; RMSE = Root-Mean-Squared Error

Table S4. The model coefficients from a Cox regression estimating the relationship between training load and risk of injury in a handball cohort (n players = 205, n injuries = 472).

Term[1,2]	HR	95% CI Lower–Upper	SE	DF	p-value
sRPE 1	0.80	0.11–5.70	0.897	11.758	0.81
sRPE 2	0.99	0.87–1.13	0.059	11.909	0.88
sRPE 3	0.77	0.01–99.10	2.259	13.435	0.91
sRPE 4	0.96	0.70–1.33	0.150	13.445	0.81
Age	0.97	0.79–1.21	0.109	456.684	0.80
Sex					
Female (Reference)	-	-	-	-	-
Male	1.13	0.781-1.641	0.189	462.46	0.51

Abbreviations: CI = Confidence Interval; df = Degrees of Freedom; HR = Hazard Ratio; SE = Standard Error; sRPE = session Rating of Perceived Exertion

[1]The frailty term for within-individual variance was significant at $p < 0.00001$

[2]The sRPE terms are the four intervals demarcated by 3 knots in the restricted cubic splines

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:

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

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.

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

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).

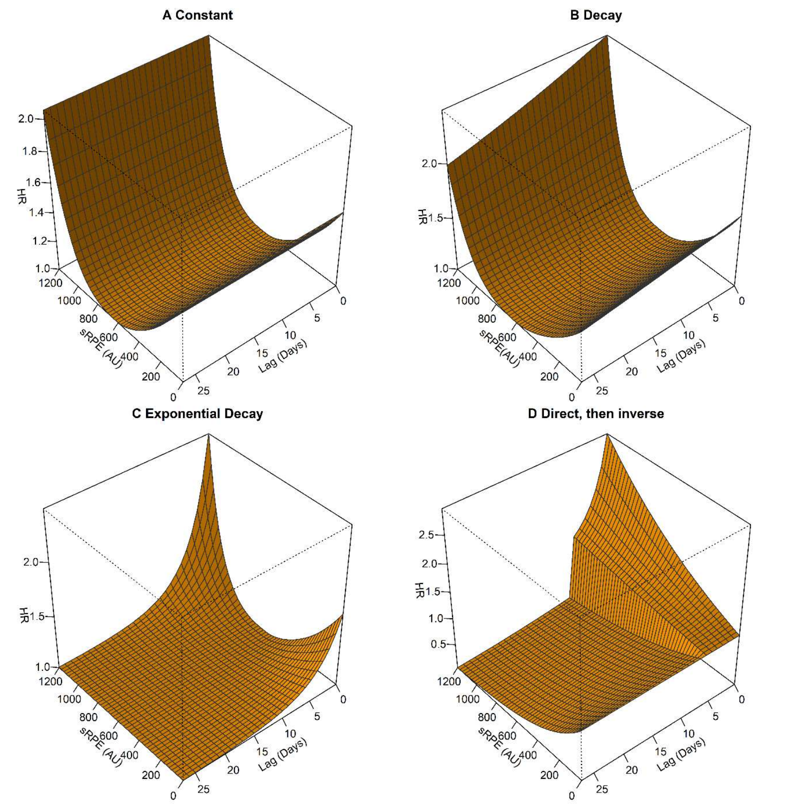

Figure 1. The four simulated relationships between absolute training load and injury risk. The relationships are a combination of the J-shaped function on the absolute training load exposure (online supplemental file 1 figure S2A) and the different functions on the time since training load was sustained (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).

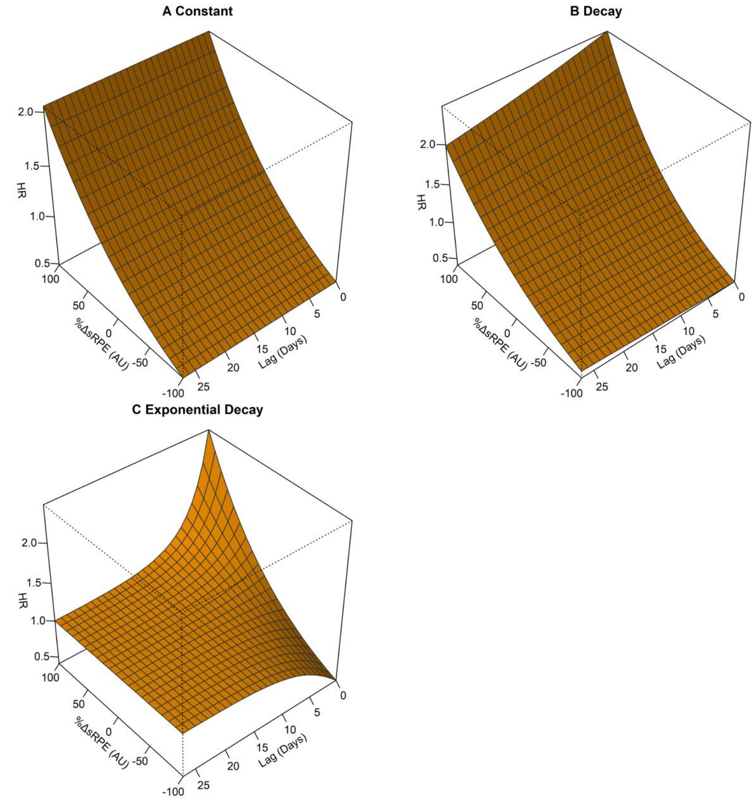

Figure 2. The three simulated relationships between relative training load and injury risk. The relationships are a combination of the linear function on the relative training load exposure (online supplemental file 1 figure S2C) and the different functions on the time since training load was sustained (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 $l = 0, \ldots, L$ 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 ($\lambda \to 0$), 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, $s(x_t, \dots, x_{t-L})$. In a distributed lag model, the effects from the lag-response relationship is modelled with the lag-response function $w(l)$:

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

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:

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

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:

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

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.

Paper IV

A new methodological approach to training load and injury risk: separating the acute from the chronic load

Lena Kristin Bache-Mathiesen[1], Thor Einar Andersen[1], Torstein Dalen-Lorentsen[1,2], Montassar Tabben[3], Karim Chamari[3], Benjamin Clarsen[1,4,5], Morten Wang Fagerland[1,6]

1. Oslo Sports Trauma Research Centre, Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway
2. Department of Smart Sensors and Microsystems, SINTEF Digital, Oslo, Norway
3. Aspetar, Orthopaedic and Sports Medicine Hospital, FIFA Medical Centre of Excellence, Doha, Qatar
4. Centre for Disease Burden, Norwegian Institute of Public Health, Bergen, Norway
5. Department of Health and Functioning, Faculty of Health and Social Sciences, Western Norway University of Applied Sciences, Bergen, Norway.
6. Oslo Centre for Biostatistics and Epidemiology, Research Support Services, Oslo University Hospital, Oslo, Norway

Correspondence to M.Sc. Lena Kristin Bache-Mathiesen, Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo 0806, Norway; l.k.bache-mathiesen@nih.no

Funding

Aspetar Orthopaedic and Sports Medicine Hospital funds the Qatar Stars League injury surveillance registry. The Oslo Sports Trauma Research Centre provided the funding for performing this study, and the Norwegian elite U-19 data collection. The Oslo Sports Trauma Research Centre has been established at the Norwegian School of Sport Sciences through generous grants from the Royal Norwegian Ministry of Culture, the South-Eastern Norway Regional Health Authority, the International Olympic Committee, the Norwegian Olympic Committee & Confederation of Sport, and Norsk Tipping AS.

Data Availability Statement

The data relevant to this study are available upon reasonable request. Statistical programming R code is available in an online GitHub repository (https://github.com/lenakba/separating-acute-from-chronic).

Ethics Approval

Study approval for the Qatar Stars League football data was obtained from the Anti-Doping Lab Qatar Institutional Review Board (E2017000252). A data sharing agreement between Aspetar Orthopaedic and Sports Medicine Hospital and Oslo Sports Trauma Research Centre was signed and approved by the Aspire Zone Foundation Institutional Review Board. For the Norwegian elite U-19 football data, study protocol was approved by the Norwegian Center for Research Data (5487), the Ethical Review Board of the Norwegian School of Sport Sciences, and the South-Eastern Norway Regional Committee for Medical and Health Research Ethics (2017/1015). Ethical principles were followed in accordance with the Declaration of Helsinki, and informed consent was obtained from all participants.

Abstract

The relationship between recent (acute) training load relative to long-term (chronic) training load may be associated with sports injury risk. We explored the potential for modelling acute and chronic loads separately to address current methodological limitations. We also determined whether there was any evidence of an interaction in the relationship between acute and chronic training loads and injury risk in football.

A men's Qatar Stars League football cohort (1 465 players, 1 977 injuries), where training load was defined as the number of minutes of activity, and a Norwegian elite U-19 football cohort (81 players, 60 injuries), where the session rating of perceived exertion (sRPE) was used. Logistic regression was run with training load on the current day (acute load) and cumulative past training load estimated by distributed lag non-linear models (chronic load) as independent variables. Injury was the outcome. An interaction between acute and chronic training load was modelled.

In both football populations, we observed that the risk of injury on the current day for different values of acute training load was highest for players with low chronic load, followed by high and then medium chronic load. The slopes varied substantially between different levels of chronic training load, indicating an interaction.

Modelling acute and chronic loads separately in regression models is a suitable statistical approach for analysing the relationship between relative training load and injury risk. Sports scientists should consider exploring interactions between acute and chronic load to improve injury prevention research.

Keywords training monitoring, load monitoring, soccer, ACWR

Introduction

Researchers attempt to identify risk factors for sports injuries to protect the athletes' health and improve sport performance (Bahr & Krosshaug, 2005). One potential, modifiable risk factor is training load. Training load is the mechanical, physiological and psychological load resultant of multiple episodes of physical activity performed by an athlete (Soligard et al., 2016). Hypotheses suggest that not only high or low training load levels may affect injury risk, but also rapid increases in recent training load relative to training load incurred in the past (Gabbett, 2016); i.e. a peak in the relative training load (Drew & Finch, 2016).

Hulin et al. (2014) introduced the Acute:Chronic Workload Ratio (ACWR) to estimate the effect of relative training load on the risk of sports injury (Blanch & Gabbett, 2016; Gabbett, 2016). In their model, the most recent training load, the acute load, is divided by the past, or chronic load. In theory, the higher the ratio – the higher the acute load relative to the chronic – the higher the risk of injury (Gabbett, 2016). In addition, a low ACWR was also proposed to increase injury risk; in other words, the suggested relationship between ACWR and injury risk was U-shaped.(Blanch & Gabbett, 2016; Gabbett, 2016) After ACWR became popular, concerns were raised on its theoretical and methodological foundations (Impellizzeri et al., 2020). Among others: the number of subjective choices involved increased risk of p-hacking (Dalen-Lorentsen et al., 2021), the time lengths for the acute and chronic periods were arbitrary (West et al., 2021), and it could not handle an acute or chronic load of 0 (Impellizzeri et al., 2020).

A core principle in the theory underlying the ACWR is that the effect of the acute load depends on the amount of chronic load. If acute load is high, it may not necessarily increase injury risk if the chronic load is also high. The aim of the ACWR was therefore to adjust the acute load to the chronic load, estimating the effect of acute load properly. This adjustment is not always successful when calculating a ratio (Impellizzeri et al., 2020). Instead, Wang et al. (2020) suggested modelling the acute load and the chronic load separately. This eliminates the risk that acute load will not be properly adjusted to the chronic load. At the time of Wang et al.'s proposal, several other challenges remained unsolved, including how to estimate the cumulative effect of past training load, the chronic load. Recent research suggests this may be solved by applying the distributed lag non-linear model (DLNM) (Bache-Mathiesen et al., 2022).

The theory that the effect of acute load on injury risk depends on the level of chronic load suggests an interaction between acute and chronic loads. Previous descriptive research has studied the association of ACWR with injury for different chronic loads (Bowen et al., 2020; Stares et al., 2018), but none have so far modelled an interaction between acute and chronic loads outside of the ACWR framework. Whether an interaction can be assessed while chronic load is modelled by DLNM is also unknown. DLNM can explore time-dependent effects, but it cannot determine what time period is considered "recent" and "past" in the context of relative training load (Bache-Mathiesen et al., 2022).

We hypothesized that training stimuli on the current day does not contribute to injury risk on that day, while the accumulated stimuli (fitness) built on past activity days *does* contribute to injury risk on the current day. In addition, if the athlete does not participate in activity on the current day, the athlete is not at risk on that day (Windt & Gabbett, 2017). We argue that the current day of activity is therefore markedly different from past activity days, and it may thus be possible to consider the current day only as the acute load, and all past observations as chronic load. Investigating whether there is evidence of such an interaction between acute and chronic loads association with injury risk may elucidate whether such interactions are worth considering in future research, and whether they are possible to model using DLNM.

When assessing the causal effect of relative training load on injury risk, one key element is the statistical description of the relationship between load and injury. The primary aim of this statistical methodology study was to investigate whether modelling acute and chronic training loads separately can be used to describe this relationship. A secondary aim was to find out whether acute and chronic loads interact in their association with injury risk in football.

Materials and methods

Participants

We analysed eight competitive seasons (2015–2022) from the men's Qatar Stars League (QSL) injury surveillance registry in football (1 465 players, 1 977 injuries, see Supplemental Table S1), and one season from a Norwegian elite U-19 football cohort (81 players [45% female], 81 injuries) described in Dalen-Lorentsen et al. (2021).

Training load definition

In the QSL data, training load was defined as the daily number of minutes in training and/or match (1 136 223 observations, 12% missing data).

In the Norwegian elite U-19 data, training load was defined as the daily number of minutes of football (training and/or match), multiplied by the player's rating of perceived exertion on a scale from 0 to 10, deriving the session Rating of Perceived Exertion (sRPE, 8 494 observations, 24% missing data) (Foster et al., 2001).

Missing data were imputed using multiple imputation (see Supplemental Figure S1–S2)(Bache-Mathiesen, Andersen, Clarsen, et al., 2021; Buuren, 2011).

Injury definition

Injuries in QSL players were recorded prospectively using the Sport Medicine Diagnostic Coding System classification (Bahr et al., 2020; Orchard et al., 2020). We recorded all injuries resulting in a player being unable to fully participate in training or match play (time-loss injuries). The player was considered injured until the team medical staff allowed full participation in training and availability for match selection. We did not record injuries that occurred outside football activities. Several steps of quality control were performed to ensure injury validity (see Supplemental file). Injuries were classified as either sudden or gradual onset.

The Norwegian elite U-19 players reported daily whether they had experienced a new health problem, with Briteback AB online survey platform, Norrköping, Sweden. If they had, a clinician contacted them for a structured interview and classified the health problem as being an injury or an illness according to the Union of European Football Associations guidelines (Hägglund et al., 2005). Only injuries were analysed in this study.

Injury definitions in both populations followed the 2006 consensus statement on epidemiological studies in football (Fuller et al., 2006).

Statistical analysis

To estimate the effect of relative training load on the risk of injury, a logistic regression model was run, with injury yes/no as the binary outcome variable.

The current day of training (Day 0) was considered the acute load and modelled as an independent variable. The relationship between the current day of training and injury risk might be non-linear (Magnusson et al., 2010), and therefore we applied restricted cubic splines (RCS) with 3 knots (Bache-Mathiesen, Andersen, Dalen-Lorentsen, et al., 2021). Due to skewed training load distributions, the knot locations were subjectively chosen based on the range of the training load observations in the QSL data (QSL model) and the Norwegian elite U-19 data (Norwegian model), respectively (Bache-Mathiesen, Andersen, Dalen-Lorentsen, et al., 2021).

The chronic load was the training performed during the previous 27 days (excluding day 0). We assumed that training load values closer to the current day may contribute more to injury risk than those distant in time (Williams et al., 2017), and that the direction of effect may also change with distance in time (Gabbett, 2016). Therefore, the cumulative effect of chronic load was modelled with a distributed lag non-linear model (DLNM)(Bache-Mathiesen et al., 2022). RCS was chosen to model the effect of the magnitude of training load (3 knots), and also the effect of the time-lag (the number of days since the training was performed, 4 knots).

An interaction term was added between the acute load (Day 0) and the DLNM-estimated chronic load (Day -1 to day -27). The main result was a visualization of the predicted probabilities of injury for acute load given different levels of chronic training load. Reference levels of chronic load was chosen by finding examples of low, medium and high chronic load in the original data (Supplemental Table S2).

Since players are only at risk of injury if they participate in an activity, days in which they did not participate in any training or match were removed from the analysis. These observations were still included in the DLNM estimation of chronic load.

The models were repeated with a random intercept term at the player level to account for the possibility that some players are more likely to suffer injuries than others (Nielsen et al., 2020). To see if a simpler approach than DLNM can be suitable, the models were also repeated using the exponentially weighted moving average (EWMA) on chronic load, same as in Williams et al. (2017).

Additional analyses were performed on the QSL data. First, the interaction model with acute and chronic minutes in activity was performed on sudden- and gradual-onset injuries, separately (Bahr et al., 2020). Second, we explored the risk of injury for various levels of minutes in activity sustained in the past, using DLNM.

Statistical analyses were performed in R version 4.2.1 with DLNM (Gasparrini, 2011), mice (Buuren, 2011), lme4 (Bates et al., 2015), and slider (Vaughan, 2021). R code is available online (Bache-Mathiesen, 2022).

Results

The QSL model showed decreased probability of injury for each minute in activity on the current day (Figure 1) with statistical significance ($p < 0.001$, Table 1) – a typical pattern when players end activity early due to injury. Players who had not participated in an activity in the last 27 days were at highest risk of injury, followed by those who spent a low number of minutes in activity (Figure 1A). Players who spent a high number of minutes in activity were at higher risk than those with medium (Figure 1A). The slopes varied considerably between different levels of minutes of activity in the past, suggesting an interaction between number of minutes in activity on the current day and cumulative number of minutes in activity the previous 27 days (Figure 1A). Of 12 interaction terms, all had narrow confidence intervals, and 4 were significant (Table 1).

A similar pattern was displayed in the Norwegian model: low chronic sRPE increased risk of injury, followed by high, with the lowest risk at medium levels of chronic sRPE (Figure 1B). Also, like the QSL model, the Norwegian elite U-19 model exhibited major changes in the slopes between the different levels of cumulative chronic sRPE, indicating an interaction (Figure 1B). However, the model failed to estimate coefficients and CIs for certain spline intervals on the chronic load (Table S3).

The relationship shape between the training load variables did not change with the addition of random effects (Figure S3), and some of the coefficients were inestimable in the mixed model. Therefore, random effects were not included in the final models.

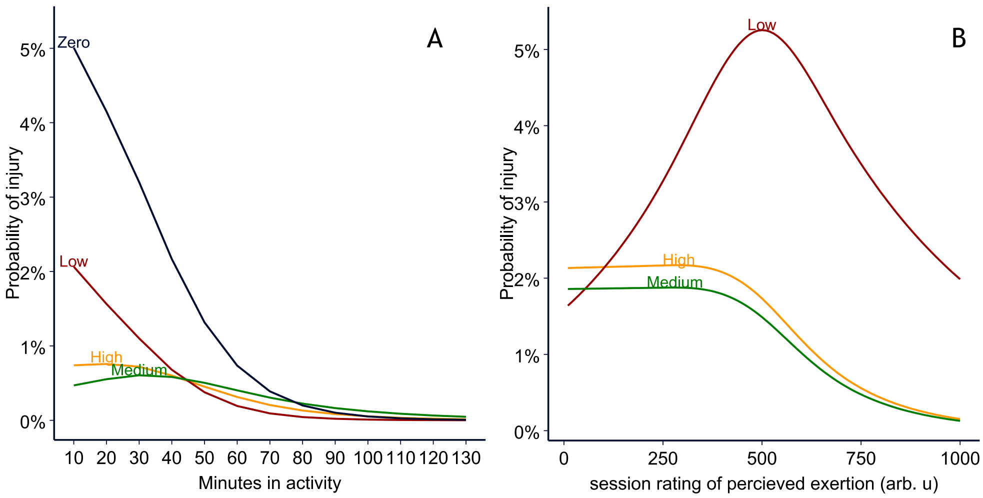

Figure 1. Probability of injury on the current day (Day 0) for each level of training load variables used in (A) Qatar Stars League model (420 329 exposure values, 1 977 injuries) and (B) Norwegian elite U-19 model (4 719 exposure values, 60 injuries). The probability is shown for zero, low, medium and high cumulative chronic training load levels. For the Qatar Stars League model: zero = 27-day sum of 0 minutes, low = 180 minutes, medium = 1435 minutes, high = 1900 minutes. For the Norwegian elite u-19 model, low = 27-day sum of 80 sRPE (near zero), medium = 7 163, high = 8 800. The exact profiles used are shown in Supplemental Table S1. The probabilities were predicted by logistic regression with an interaction term between the acute load and the cumulative chronic load. Only days in which the players were at risk were analysed (acute load ≠ 0). Arb. u = arbitrary units.

The EWMA models failed to discover an effect of chronic training load on injury risk, neither in the QSL model, nor in the Norwegian model, and did not display signs of an interaction (Figure S4). The relationship shape between acute load and injury risk was different from the DLNM models, showing a peak around 60 minutes for QSL (Figure S4A), and an exponential increase in risk for the Norwegian model.

Table 1. QSL model coefficients for a logistic regression with injury as the outcome and minutes in activity on the current day (acute), and past minutes in activity (chronic) as independent variables.

Term[123]	OR	SE	Lower CI	Upper CI	P
Intercept	0.067	0.329	0.034	0.131	<0.001
Acute minutes in activity 1	0.950	0.007	0.936	0.964	<0.001
Acute minutes in activity 2	1.144	0.017	1.105	1.185	<0.001
Chronic minutes in activity W1 F1	2.166	0.252	1.282	3.659	0.006
Chronic minutes in activity W1 F2	0.455	0.129	0.348	0.595	<0.001
Chronic minutes in activity W1 F3	1.285	0.11	1.030	1.602	0.027
Chronic minutes in activity W2 F1	0.156	0.374	0.075	0.324	<0.001
Chronic minutes in activity W2 F2	6.112	0.191	4.207	8.881	<0.001
Chronic minutes in activity W2 F3	0.841	0.181	0.590	1.200	0.340
Chronic minutes in activity W3 F1	3.252	0.623	0.952	11.109	0.060
Chronic minutes in activity W3 F2	0.578	0.363	0.279	1.198	0.137
Chronic minutes in activity W3 F3	0.673	0.281	0.388	1.168	0.159
Chronic minutes in activity W4 F1	6.432	1.228	0.578	71.55	0.130
Chronic minutes in activity W4 F2	0.319	0.642	0.090	1.126	0.076
Chronic minutes in activity W4 F3	0.404	0.573	0.130	1.256	0.116
Interaction (Acute*Chronic minutes W1 F1)	0.998	0.003	0.991	1.006	0.642
Interaction (Acute*Chronic minutes W1 F2)	1.002	0.002	0.998	1.006	0.429
Interaction (Acute*Chronic minutes W1 F3)	1.000	0.001	0.998	1.003	0.844
Interaction (Acute*Chronic minutes W2 F1)	1.020	0.004	1.012	1.028	<0.001
Interaction (Acute*Chronic minutes W2 F2)	0.978	0.002	0.974	0.982	<0.001
Interaction (Acute*Chronic minutes W2 F3)	1.004	0.002	1.001	1.008	0.020
Interaction (Acute*Chronic minutes W3 F1)	0.993	0.006	0.982	1.005	0.243
Interaction (Acute*Chronic minutes W3 F2)	1.010	0.003	1.003	1.017	0.009
Interaction (Acute*Chronic minutes W3 F3)	1.003	0.003	0.997	1.009	0.340
Interaction (Acute*Chronic minutes W4 F1)	0.996	0.009	0.978	1.015	0.678
Interaction (Acute*Chronic minutes W4 F2)	1.005	0.005	0.995	1.015	0.311
Interaction (Acute*Chronic minutes W4 F3)	1.007	0.005	0.997	1.016	0.154

Abbreviations: CI = 95% Confidence Interval, OR = Odds Ratio, QSL = Qatar Stars League, SE = Standard Error
[1]All variables were modelled with splines (420 329 exposure values, 1 977 injuries), and terms represent one of multiple intervals demarcated by knots
[2]The DLNM models a cross-product of the number of minutes in activity (the F-function) and the lag time in which the activity was performed (the W-function). Since F was modelled with 3 knots, and W with 4, the result is a 3*4 permutation of intervals

QSL players were at higher risk of a sudden onset than a gradual onset injury (Figure S5). Signs of an interaction between minutes in activity on the current day and cumulative minutes in activity in the past were present in both acute and overuse injuries (Figure S5).

In the QSL population, activities performed on the day before the current day contributed most to the risk of injury on the current day (OR = 1.1 for 60 minutes of activity, 95% confidence interval (CI) = 1.05–1.18, Figure 2). The risk declined exponentially the more distant in time the activity was performed, ending at approximately OR = 1.02 (CI = 1.01-1.04) for 60 minutes of activity performed 19 to 22 days prior to the current day. At 27 days prior to the current day, the CI overlapped with 1 (OR = 1.02, 0.98–1.07). A low number of minutes in activity (10–40 minutes) on a day in the past substantially increased risk of injury for the current day, a high number (90–120 minutes) moderately increased risk, and a medium number (40–80 minutes) slightly increased risk, regardless of whether the activity was performed 1 day prior to the current day, 10 days prior, or 27 days prior (Figure 2B–D).

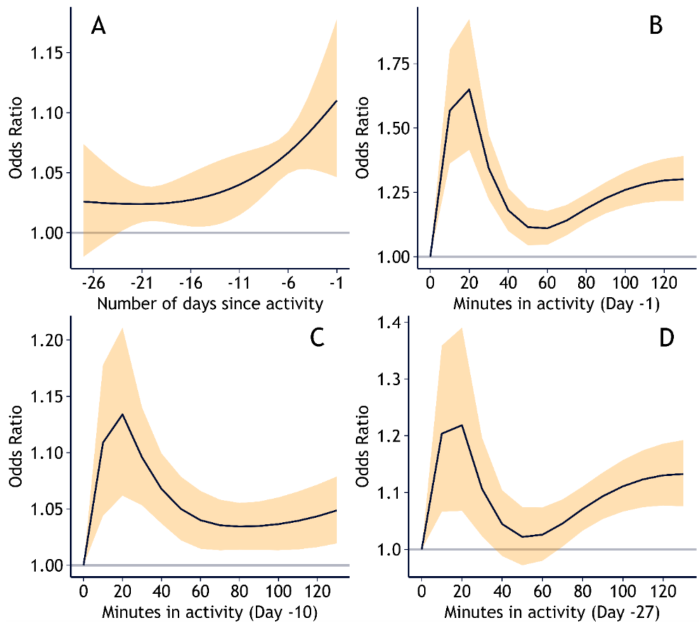

Figure 2. Risk profiles of chronic load, measured by the minutes of activity performed by Qatar Stars League players (1 136 223 exposure values, 1 977 injuries). Figure A shows the risk of 60 minutes of activity for each day in the past. -1 is the risk of injury if 60 minutes of activity occurred the day prior to the current day, and -27 is the risk if 60 minutes of activity occurred 27 days before the current day. Figures B, C, and D shows how the risk of injury changes for each level of minutes in activity if the activity occurred (B) 1 day prior to the current day, (C) 10 days prior to the current day, (D) 27 days prior to the current day. Note, that Y-axes for B–D are not on the same scale, to better show the relationship shape. Coefficients with 95% confidence intervals (yellow bands) were estimated by a logistic regression model with injury as the outcome and a distributed lag non-linear model of the chronic minutes in activity as the independent variable. No other terms were included.

Discussion

This is the first study to explore the potential of modelling acute and chronic training loads separately to estimate the effect of relative training load on injury risk in sport. The method successfully discovered a relationship between the minutes in activity on the current day and the probability of injury in a Qatar Stars League football (soccer) population, and properly adjusted for the cumulative effect of minutes in activity the previous 27 days. Signs of a relationship between internal training load (sRPE) on the current day and injury risk could also be gleaned in a Norwegian elite U-19 football population, although with high uncertainty due to a much smaller sample size. In future observational studies of causal inference, this statistical approach can be used to determine whether relative training load affects injury risk, given that confounding and other considerations for causal inference have been properly addressed.

This study also investigated, for the first time, whether there was an interaction between acute and chronic training loads, where "acute" was defined as the current day of activity. Clear evidence of an interaction was found, as in both the QSL model and the Norwegian model, the relationship slopes for acute training load varied considerably for different levels of chronic training load. This demonstrates that interactions between time periods can be modelled with the DLNM approach.

Modelling acute and chronic loads separately

The QSL model indicated decreased injury risk for each minute spent in activity on the current day ($p < 0.001$). The Norwegian model displayed a similar trend, although non-significant ($p > 0.05$), and injury risk increased if chronic internal load (cumulative past sRPE) was low. We suspect that players who ended activity due to injury skewed the models toward decreased risk with increased exposure. This effect was amplified in the QSL population, which only included time-loss injuries and time in exposure – no measure of the training intensity. This is a general and – yet – unsolved challenge for studies that aim to estimate the effect of training load on injury risk.

Both the QSL and the Norwegian models displayed variation in injury risk given different levels of cumulative chronic load. Low chronic training load had highest risk, followed by high chronic load, then the medium load with the lowest risk. In summary, modelling the acute and chronic load separately successfully estimated the effect of acute load adjusted for the effect of chronic load. In addition, having zero chronic load the last four weeks (a month without football) showed the highest risk of injury in the QSL model. This could not have been discovered if we had used any form of ratio, as the denominator would be 0 (Curran-Everett, 2013). Lastly, while using the ACWR would require choosing among multiple ways of calculation (Drew et al., 2017; Gabbett et al., 2019; Williams et al., 2017), the current approach required few such choices, and reduced the risk of multiple testing issues.

The EWMA model failed to discover a relationship between chronic load and injury risk and could not separate the effects of different chronic training load levels. Given the large sample size of 1 136 223 observations and 1 977 injuries of the QSL population, we speculate whether EWMA could estimate the effects at all, even in a larger study.

Interaction between acute and chronic loads in football

Interestingly, the slopes of the effect of chronic load on injury risk varied considerably in the two football populations. High and medium chronic load slowly declined in risk for each level of acute load, while low chronic load declined rapidly (Figure 1A). In the Norwegian model, low chronic load both increased and decreased risk at different levels of acute load (Figure 1B). In the QSL population, the interaction was also present when stratified on acute and overuse injuries. We recommend that future training load and injury risk studies consider and explicitly model these interactions, to improve injury prevention research.

Choosing time periods for acute and chronic load

A consistent challenge with traditional methods of estimating relative training load's effect on injury risk, whether it be the acute:chronic workload ratio or other methods, is choosing the time periods for acute and chronic load (Coyne et al., 2022; West et al., 2021). If recent days of training increase risk, and past days of training reduce risk, at what point in time does this change happen? Subjectively deducing the cut-off may be arbitrary (Coyne et al., 2022), cut-offs based on previous research may not be sport-specific (Impellizzeri et al., 2020), and data-driven approaches risk multiple testing issues and reduced comparability (Carey et al., 2017).

Modelling the effects of chronic load using DLNM allowed exploration of the time-lag structure. In the QSL population, the risk of injury declined exponentially the further distant in time the activity was performed: OR = 1.10 (CI = 1.05–1.18) for 60 minutes of activity performed the day prior to the current day, OR = 1.06 (CI = 1.05–1.08) for the same amount performed 6 days prior to the current day, ending at OR = 1.02 (CI = 1.01–1.04) for 60 minutes of activity performed 21 days prior to the current day. Furthermore, a low number (10–40) or a high number (90–120) of minutes in activity on a day in the past both increased risk of injury on the current day, while a medium number (40–80 minutes) decreased risk in comparison. This reflected the pattern seen in Figure 1, and it fits the hypotheses that both too much and too little training may increase risk of injury (Gamble, 2013). The DLNM approach can explore at what point in time in the past the effect of chronic load changes (if it changes).

We hypothesized that the current day (Day 0) has special properties compared to past days of training load exposure, which allows it to be modelled separately without the concern that it may be too similar to concurrent days.

On the current day, injury risk increases with sheer exposure to the physical activity itself. Players cannot sustain an injury if they do not participate in an activity (Gabbett, 2016). On the other hand, if players did not participate in an activity on certain days in the past, those days would still contribute to the cumulative effect of past training load. Thus, the effect of a training load value of 0 changes drastically if it is on the current day versus past training load days.

Hypotheses suggest that both high and low levels of training load may increase injury-risk (Gamble, 2013). Too little training will not build enough fitness for the tissue to tolerate upcoming training load levels. Too much training may potentially damage the tissue, and the tissue may not be able to regenerate in time for the next training or match-play exposure. These hypotheses pertain mostly to past training load. On the current day, the player enters with fitness and fatigue resultant of the past. The adaptations built during the current day of activity will not likely come into play until later (that day or during the successive days). The fatigue, will, however affect the current activity and day. Hence, the shape of the relationship between training load and injury risk (linear, or various non-linear), may depend on whether the event was in the past, or on the current day.

In a real-time setting, the current and future days of training or match-play load are the most modifiable. One cannot change training load that happened in the past. Coaching staff, medical staff and players (athletes) are interested in the risk of injury on the current day and future days –

given the training that was done in the past. Studies interested in causal inference and developing load management programs should take this into consideration when choosing time periods for acute and chronic loads.

Future perspectives

We showed the potential of modelling acute and chronic training loads separately. DLNM is a flexible approach to handling the complexity of chronic load; moreover, the DLNM R-package is free and available online. The R-package was, however, developed for epidemiological questions, and not yet adapted to interactions or stratified analysis. Consequently, its prediction functionalities could not be used, which barred us from exploring effects in the time-lag structure of the interaction model. Furthermore, predictions for different levels of chronic load could not be set to a desired level and therefore, examples were selected from the original data. Future research is needed in implementation of DLNM for the niche of training load.

While this study focused on football, we believe the proposed method is flexible enough to handle sport-specific circumstances, such as tapering (Murach & Bagley, 2015), and can be used in both individual and team sports, warranting interesting studies.

Limitations

Limitations of this study were: (i) the QSL data only had minutes of activity, and no other training load variables or variable describing the intensity of the activity; (ii) the Norwegian elite U-19 data had only sRPE – the player's psychological perception of the training exertion and the duration of the activity. The sRPE has recently been critiqued (Passfield et al., 2022), and different groups of players can perceive the same physiological stimuli differently (Impellizzeri et al., 2004); the Norwegian elite U-19 sRPE responses were above other football populations (Chamari et al., 2012; Rabbani et al., 2019). In this regard, training load is a multidimensional construct, and ideally, both internal and external training loads should be used (Bourdon et al., 2017).

We considered only the current day to be the acute load. We therefore could not uncover whether the effects of relative training load existed more distant in the past; for instance, if the training performed three days ago relative to the training performed six days ago (or other time variations) had an association with injury risk. In addition, due to multicollinearity, confidence intervals around predictions in Figure 1 could not be estimated.

Conclusion

To assess the effect of recent (acute) training load relative to past (chronic) training load on injury risk, a ratio has traditionally been calculated. Ratios have a number of challenges, including how to handle chronic loads of 0. Modelling the acute and the chronic load separately is intuitive and potentially a simple solution to this problem. When using this method, the acute load adjusts for the level of chronic load without calculating a ratio. Furthermore, signs of an interaction between acute and chronic training load were present in both football populations studied. Scientists in the field of training load and injury risk should consider and model these interactions to improve injury prevention research.

Acknowledgements

Thank you to Prof. Roald Bahr and Prof. Marco Cardinale for arranging the Qatar research internship. Thanks also to Neshi Arif and Anna Kochergina for contributing to facilitating this research. We also thank the Aspetar Orthopaedic and Sports Medicine Hospital for the collaboration.

Disclosure of interest

The authors report there are no competing interests to declare.

References

Bache-Mathiesen, L. K. (2022). *Modelling the acute and chronic loads separately: R code repository*. In (Version 1.0) [R code repository]. https://github.com/lenakba/separating-acute-from-chronic

Bache-Mathiesen, L. K., Andersen, T. E., Clarsen, B., & Fagerland, M. W. (2021). Handling and reporting missing data in training load and injury risk research. *Science and Medicine in Football*, 1-13. https://doi.org/10.1080/24733938.2021.1998587

Bache-Mathiesen, L. K., Andersen, T. E., Dalen-Lorentsen, T., Clarsen, B., & Fagerland, M. W. (2021). Not straightforward: modelling non-linearity in training load and injury research. *BMJ Open Sport & Exercise Medicine*, *7*(3), e001119. https://doi.org/10.1136/bmjsem-2021-001119

Bache-Mathiesen, L. K., Andersen, T. E., Dalen-Lorentsen, T., Clarsen, B., & Fagerland, M. W. (2022). Assessing the cumulative effect of long-term training load on the risk of injury in team sports. *BMJ Open Sport & Exercise Medicine*, *8*(2), e001342. https://doi.org/10.1136/bmjsem-2022-001342

Bahr, R., Clarsen, B., Derman, W., Dvorak, J., Emery, C. A., Finch, C. F., Hägglund, M., Junge, A., & Kemp, S. (2020). International Olympic Committee Consensus Statement: Methods for Recording and Reporting of Epidemiological Data on Injury and Illness in Sports 2020 (Including the STROBE Extension for Sports Injury and Illness Surveillance (STROBE-SIIS)). *Orthopaedic journal of sports medicine*, *8*(2). https://doi.org/10.1177/2325967120902908

Bahr, R., & Krosshaug, T. (2005). Understanding injury mechanisms: a key component of preventing injuries in sport. *British journal of sports medicine*, *39*(6), 324-329. https://doi.org/10.1136/bjsm.2005.018341

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67(1)*, 1-48. https://doi.org/10.18637/jss.v067.i01

Blanch, P., & Gabbett, T. J. (2016). 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*, *50*(8), 471-475. https://doi.org/10.1136/bjsports-2015-095445

Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M. C., Gabbett, T. J., Coutts, A. J., Burgess, D. J., & Gregson, W. (2017). Monitoring athlete training loads: consensus statement. *International journal of sports physiology and performance*, *12*(s2), S2-161-S162-170. https://doi.org/10.1123/IJSPP.2017-0208

Bowen, L., Gross, A. S., Gimpel, M., Bruce-Low, S., & Li, F.-X. (2020). Spikes in acute: chronic workload ratio (ACWR) associated with a 5–7 times greater injury rate in English Premier League football players: a comprehensive 3-year study. *British journal of sports medicine*, *54*, 731-738. https://doi.org/10.1136/bjsports-2018-099422

Buuren, S. v. G.-O., Karin. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1-67. https://doi.org/10.18637/jss.v045.i03

Carey, D. L., Blanch, P., Ong, K.-L., Crossley, K. M., Crow, J., & Morris, M. E. (2017). Training loads and injury risk in Australian football—differing acute: chronic workload ratios influence match injury risk. *British journal of sports medicine*, *51*(16), 1215-1220. https://doi.org/10.1136/bjsports-2016-096309

Chamari, K., Haddad, M., Wong, D. P., Dellal, A., & Chaouachi, A. (2012). Injury rates in professional soccer players during Ramadan. *Journal of sports sciences*, *30*(sup1), S93-S102. https://doi.org/10.1080/02640414.2012.696674

Coyne, J. O. C., Coutts, A. J., Newton, R. U., & Haff, G. G. (2022). The Current State of Subjective Training Load Monitoring: Follow-Up and Future Directions. *Sports Medicine - Open*, *8*(1), 53. https://doi.org/10.1186/s40798-022-00433-y

Curran-Everett, D. (2013). Explorations in statistics: the analysis of ratios and normalized data. *Advances in physiology education*, *37*(3), 213-219. https://doi.org/10.1152/advan.00053.2013

Dalen-Lorentsen, T., Andersen, T. E., Bjørneboe, J., Vagle, M., Martin, K. N., Kleppen, M., Fagerland, M. W., & Clarsen, B. (2021). A Cherry, Ripe for Picking: The Relationship Between the Acute-Chronic Workload Ratio and Health Problems. *Journal of Orthopaedic & Sports Physical Therapy*, *51*(4), 162-173. https://doi.org/10.2519/jospt.2021.9893

Drew, M. K., Blanch, P., Purdam, C., & Gabbett, T. J. (2017). 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*, *51*(7), 618-619. https://doi.org/10.1136/bjsports-2016-096609

Drew, M. K., & Finch, C. F. (2016). The relationship between training load and injury, illness and soreness: a systematic and literature review. *Sports Medicine*, *46*(6), 861-883. https://doi.org/10.1007/s40279-015-0459-8

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A New Approach to Monitoring Exercise Training. *Journal of strength and conditioning research*, *15*(1), 109-115. https://europepmc.org/article/med/11708692

Fuller, C. W., Ekstrand, J., Junge, A., Andersen, T. E., Bahr, R., Dvorak, J., Hägglund, M., McCrory, P., & Meeuwisse, W. H. (2006). Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Scandinavian journal of medicine & science in sports*, *16*(2), 83-92. https://doi.org/10.1111/j.1600-0838.2006.00528.x

Gabbett, T. J. (2016). The training—injury prevention paradox: should athletes be training smarter and harder? *British journal of sports medicine*, *50*(5), 273-280. https://doi.org/10.1136/bjsports-2015-095788

Gabbett, T. J., Hulin, B., Blanch, P., Chapman, P., & Bailey, D. (2019). To couple or not to couple? For acute: chronic workload ratios and injury risk, does it really matter? *International journal of sports medicine*, *40*(09), 597-600. https://doi.org/10.1055/a-0955-5589

Gamble, P. (2013). Reducing injury in elite sport-is simply restricting workloads really the answer? [Review]. *New Zealand Journal of Sports Medicine*, *40*, 36-38. https://sportsmedicine.co.nz/wp-content/uploads/2016/02/NZJSM-Vol-40-Issue-1.pdf#page=38

Gasparrini, A. (2011). Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software*, *43*(8), 1. https://doi.org/10.18637/jss.v043.i08

Hulin, B. T., Gabbett, T. J., Blanch, P., Chapman, P., Bailey, D., & Orchard, J. W. (2014). Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *British journal of sports medicine*, *48*(8), 708-712. https://doi.org/10.1136/bjsports-2013-092524

Hägglund, M., Waldén, M., Bahr, R., & Ekstrand, J. (2005). Methods for epidemiological study of injuries to professional football players: developing the UEFA model. *British journal of sports medicine*, *39*(6), 340-346. https://doi.org/10.1136/bjsm.2005.018267

Impellizzeri, F. M., Rampinini, E., Coutts, A. J., Sassi, A., & Marcora, S. M. (2004). Use of RPE-based training load in soccer. *Medicine & Science in Sports & Exercise*, *36*(6), 1042-1047. https://doi.org/10.1249/01.MSS.0000128199.23901.2F

Impellizzeri, F. M., Tenan, M. S., Kempton, T., Novak, A., & Coutts, A. J. (2020). Acute: Chronic Workload Ratio: Conceptual Issues and Fundamental Pitfalls. *International journal of sports physiology and performance*, *15*(6), 907-913. https://doi.org/10.1123/ijspp.2019-0864

Magnusson, S. P., Langberg, H., & Kjaer, M. (2010). The pathogenesis of tendinopathy: balancing the response to loading. *Nature Reviews Rheumatology*, *6*(5), 262-268. https://doi.org/10.1038/nrrheum.2010.43

Murach, K. A., & Bagley, J. R. (2015). Less Is More: The Physiological Basis for Tapering in Endurance, Strength, and Power Athletes. *Sports*, *3*(3), 209-218. https://doi.org/doi:10.3390/sports3030209

Nielsen, R. O., Shrier, I., Casals, M., Nettel-Aguirre, A., Møller, M., Bolling, C., Bittencourt, N. F., Clarsen, B., Wedderkopp, N., & Soligard, T. (2020). Statement on Methods in Sport Injury Research From the First METHODS MATTER Meeting, Copenhagen, 2019. *Journal of Orthopaedic & Sports Physical Therapy*, *50*(5), 226-233. https://doi.org/10.1136/bjsports-2019-101323

Orchard, J. W., Meeuwisse, W., Derman, W., Hägglund, M., Soligard, T., Schwellnus, M., & Bahr, R. (2020). Sport Medicine Diagnostic Coding System (SMDCS) and the Orchard Sports Injury and Illness Classification System (OSIICS): revised 2020 consensus versions. *British journal of sports medicine*, *54*(7), 397-401. https://doi.org/10.1136/bjsports-2019-101921

Passfield, L., Murias, J. M., Sacchetti, M., & Nicolò, A. (2022). Validity of the Training-Load Concept. *International journal of sports physiology and performance*, 1-8. https://doi.org/10.1123/ijspp.2021-0536

Rabbani, A., Clemente, F. M., Kargarfard, M., & Chamari, K. (2019). Match Fatigue Time-Course Assessment Over Four Days: Usefulness of the Hooper Index and Heart Rate Variability in Professional Soccer Players [Original Research]. *Frontiers in physiology*, *10*. https://doi.org/10.3389/fphys.2019.00109

Soligard, T., Schwellnus, M., Alonso, J.-M., Bahr, R., Clarsen, B., Dijkstra, H. P., Gabbett, T., Gleeson, M., Hägglund, M., & Hutchinson, M. R. (2016). How much is too much?(Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *British journal of sports medicine*, *50*(17), 1030-1041. https://doi.org/10.1136/bjsports-2016-096581

Stares, J., Dawson, B., Peeling, P., Heasman, J., Rogalski, B., Drew, M., Colby, M., Dupont, G., & Lester, L. (2018). Identifying high risk loading conditions for in-season injury in elite Australian football players. *Journal of science and medicine in sport*, *21*(1), 46-51. https://doi.org/10.1016/j.jsams.2017.05.012

Vaughan, D. (2021). *slider: Sliding Window Functions*. In (Version 0.2.1) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=slider

Wang, C., Vargas, J. T., Stokes, T., Steele, R., & Shrier, I. (2020). Analyzing Activity and Injury: Lessons Learned from the Acute: Chronic Workload Ratio. *Sports Medicine*, 1-12. https://doi.org/10.1007/s40279-020-01280-1

West, S., Williams, S., Cazzola, D., Cross, M., Kemp, S., & Stokes, K. (2021). 119 The acute:chronic workload ratio: why one size does not fit all. *British journal of sports medicine*, *55*(Suppl 1), A48-A48. https://doi.org/10.1136/bjsports-2021-IOC.111

Williams, S., West, S., Cross, M. J., & Stokes, K. A. (2017). Better way to determine the acute: chronic workload ratio? *British journal of sports medicine*, *51*(3), 209-210. https://doi.org/10.1136/bjsports-2016-096589

Windt, J., & Gabbett, T. J. (2017). How do training and competition workloads relate to injury? The workload—injury aetiology model. *British journal of sports medicine*, *51*(5), 428-435. https://doi.org/10.1136/bjsports-2016-096040

Supplementary

Injury validation in Qatar Stars League registry

The team physician in each club was in charge of collecting the data, using standardized tools. We distributed a study manual outlining the details of data collection to the contact person before the team's enrollment into the study. We also organized demonstration sessions every time a new team physician joined the program. We recorded data using a custom-made Microsoft Office Excel® file (Microsoft Corporation, Readmon, WA, USA) for quick data entry, using pull-down menus to classify each injury based on the Sport Medicine Diagnostic Coding System. Injury cards were also provided in Microsoft Office Word® (Microsoft Corporation, Readmon, WA, USA) to assist clinicians in taking notes during daily clinical activity, prior to entry into the master data file. We asked the clubs to submit their data every month by email. Data quality control was done on a monthly basis to validate the data.

Tables

Table S1. Characteristics of 1 465 Qatar Stars League players for the 3 365 studied player' seasons.

Characteristic[1]	**Mean (SD)**
Age (n = 564)	25 (5)
Height (n = 535)	174 (21)
Weight (n = 548)	71 (16)
Player position (n = 725)[2]	**N (%)**
Defenders	231 (32%)
Goal Keepers	81 (11%)
Midfielders	316 (44%)
Strikers	97 (13%)

[1]Variables had missing data, and descriptives are calculated on observed values (n).
[2]One player could change positions across multiple seasons, and therefore be included multiple times in the calculation

Table S2. Chronic load profiles used as reference values in Figure 1 (main article), Figure S3 and Figure S5, from the day before the current day (-1) to 27 days prior to the current day (-27).

	Qatar Stars League[1]				Norwegian elite U-19[2]		
Day	**Zero**	**Low**	**Medium[3]**	**High[4]**	**Low**	**Medium[3]**	**High[4]**
-1	0	60	90	45	80	480	720
-2	0	60	27	45	0	0	630
-3	0	60	79	90	0	720	540
-4	0	0	60	80	0	588	1260
-5	0	0	30	80	0	120	0
-6	0	0	63	80	0	0	560
-7	0	0	30	90	0	450	0
-8	0	0	63	140	0	30	0
-9	0	0	60	105	0	0	1230
-10	0	0	11	70	0	540	0
-11	0	0	78	40	0	900	810
-12	0	0	15	15	0	390	0
-13	0	0	77	45	0	90	0
-14	0	0	13	45	0	240	0
-15	0	0	78	90	0	370	320
-16	0	0	75	90	0	30	0
-17	0	0	0	90	0	360	0
-18	0	0	0	90	0	60	0
-19	0	0	70	90	0	0	0
-20	0	0	70	90	0	540	0
-21	0	0	70	45	0	55	630
-22	0	0	26	90	0	0	360
-23	0	0	70	30	0	0	0
-24	0	0	70	45	0	0	960
-25	0	0	70	45	0	30	360
-26	0	0	70	90	0	540	0
-27	0	0	70	45	0	630	420
Total	0	180	1435	1900	80	7163	8800

[1]Measured in minutes in activity
[2]Measured in session Rating of Perceived Exertion (sRPE) in arbitrary units
[3]The total sum was the median in the corresponding dataset
[4]The total sum was the 75% quantile in the corresponding dataset

Table S3. Model coefficients for a logistic regression with injury as the outcome and sRPE on the current day (acute), and past sRPE (chronic) as independent variables in the Norwegian elite U-19 data.

Term[123]	OR	SE	Lower CI	Upper CI	p
Intercept	0.035	1.110	0.004	0.308	0.003
Acute sRPE	1.001	0.003	0.996	1.006	0.656
Acute sRPE	0.997	0.002	0.992	1.001	0.177
Chronic sRPE W1 F1	0.111	1.055	0.014	0.883	0.038
Chronic sRPE W1 F2	0.972	0.660	0.266	3.544	0.965
Chronic sRPE W1 F3	2.661	0.638	0.758	9.343	0.126
Chronic sRPE W2 F1	369558.600	4.787	30.843	4.43E+09	0.007
Chronic sRPE W2 F2	0.122	2.538	0.001	17.66	0.407
Chronic sRPE W2 F3	0.230	2.724	0.001	48.581	0.589
Chronic sRPE W3 F1	0.000	15.939	0.000	390.613	0.108
Chronic sRPE W3 F2	13.162	6.383	0.000	3647113	0.686
Chronic sRPE W3 F3	4.529	6.798	0.000	2924533	0.824
Chronic sRPE W4 F1	0.000	33.76	0.000	0.324	0.046
Chronic sRPE W4 F2	22218.120	13.56	0.000	8.02E+15	0.461
Chronic sRPE W4 F3	92.306	15.116	0.000	8.11E+14	0.765
Interaction (Acute*Chronic sRPE W1 F1)	1.005	0.002	1.001	1.009	0.016
Interaction (Acute*Chronic sRPE W1 F2)	1.000	0.001	0.997	1.002	0.866
Interaction (Acute*Chronic sRPE W1 F3)	0.999	0.001	0.996	1.001	0.259
Interaction (Acute*Chronic sRPE W2 F1)	0.971	0.009	0.954	0.988	0.001
Interaction (Acute*Chronic sRPE W2 F2)	1.005	0.005	0.996	1.014	0.310
Interaction (Acute*Chronic sRPE W2 F3)	0.999	0.005	0.990	1.009	0.900
Interaction (Acute*Chronic sRPE W3 F1)	1.056	0.026	1.003	1.111	0.038
Interaction (Acute*Chronic sRPE W3 F2)	0.989	0.014	0.962	1.016	0.418
Interaction (Acute*Chronic sRPE W3 F3)	1.008	0.012	0.984	1.033	0.500
Interaction (Acute*Chronic sRPE W4 F1)	1.161	0.057	1.039	1.298	0.009
Interaction (Acute*Chronic sRPE W4 F2)	0.967	0.030	0.912	1.025	0.262
Interaction (Acute*Chronic sRPE W4 F3)	1.017	0.027	0.964	1.074	0.535

Abbreviations: CI = 95% Confidence Interval, OR = Odds Ratio, SE = Standard Error, sRPE = session Rating of Perceived Exertion in arbitrary units

[1]All variables were modelled with splines, and terms represent one of multiple intervals demarcated by knots

[2]The DLNM models a crossproduct of the number of minutes in activity (the F-function) and the lag time in which the activity was performed (the W-function). Since F was modelled with 3 knots, and W with 4, the result is a 3*4 permutation of intervals

Figures

Figure S1. Illustration of the modelling process in the framework of multiple imputation. The imputation was performed in accordance with recommendations in "Flexible Imputation of Missing Data, Second Edition" by Stef van Buuren (Van Buuren, 2018a), also available online (Van Buuren, 2018b). Missing time in activity in minutes, and sRPE values, were predicted and imputed using predictive mean matching (Barzi & Woodward, 2004), which has previously been shown to be a valid approach for count data (Van Buuren, 2018a). For the minutes in activity, a poisson regression imputation was compared with the PMM with validation plots, before choosing PMM. All non-derived variables were used to predict imputed values, including age, sex, player position, type of training activity, among others. The response variable, injury, was also used to predict imputed values (Moons et al., 2006), but was not itself imputed before analysis (Peters et al., 2012). The number of imputed datasets was five, which is recommended in most cases (Van Buuren section 2.8). The imputation was validated by comparing the distribution of the imputed versus the original data (see Figure S2). Five models were fitted and pooled using Ruben's Rules for the final models (results in Table 1 and Table 2, main article).

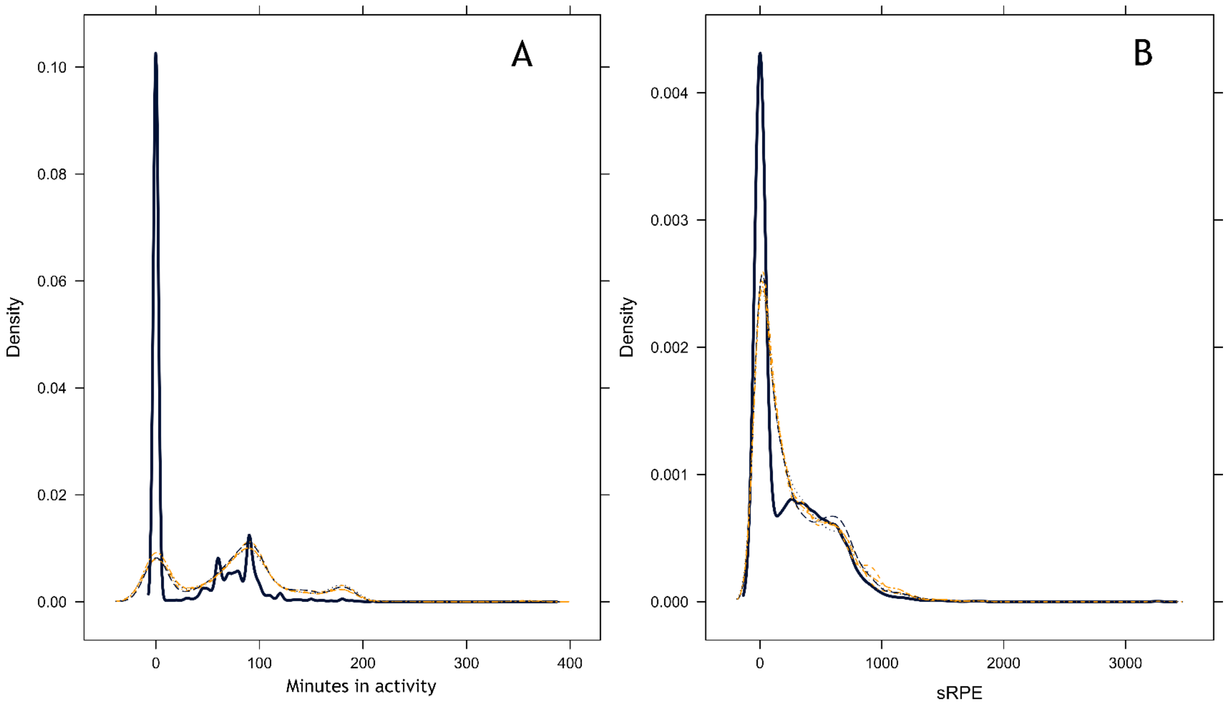

Figure S2. Distribution of original data values (blue) compared to imputed values from five imputed datasets (yellow) for (A) daily minutes in activity in a Qatar Stars League football population, and (B) daily session Rating of Perceived Exertion (sRPE) measured in arbitrary units in a Norwegian elite U-19 football cohort. The mismatch between the distribution of imputed data and original data in (A) is expected. Although 12% of the Qatar Stars League exposure observations were missing, on days that players suffered an injury, the missing rate was 36%. The missing mechanism was therefore missing at random, and missing probability increased if injury = yes. Since players are unlikely to be injured on days with no activity (exposure = 0), one would expect the imputed distribution to skew less towards 0 than the original data.

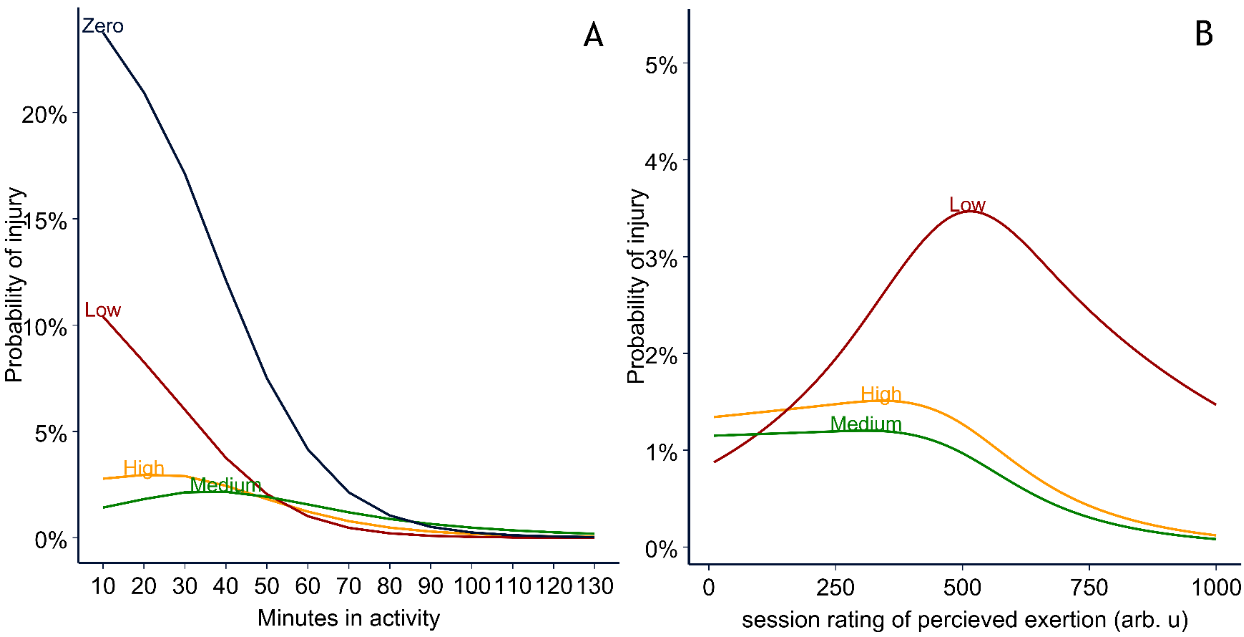

Figure S3. Probability of injury on the current day (Day 0) predicted by logistic regression models with random effects. Shown for each level of training load variables used in (A) Qatar Stars League model (420 329 exposure values, 1 977 injuries) and (B) Norwegian elite U-19 model (4 719 exposure values, 60 injuries). The probability is shown for zero, low, medium and high cumulative chronic training load levels. For the Qatar Stars League model, the sums chosen were: zero = 27-day sum of 0 minutes, low = 180 minutes, medium = 1 435 minutes, high = 1 900 minutes. For the Norwegian elite u-19 model, low = 80, medium = 7 163, high = 8 800. The exact profiles used are shown in Table S1. Arb. u = arbitrary units.

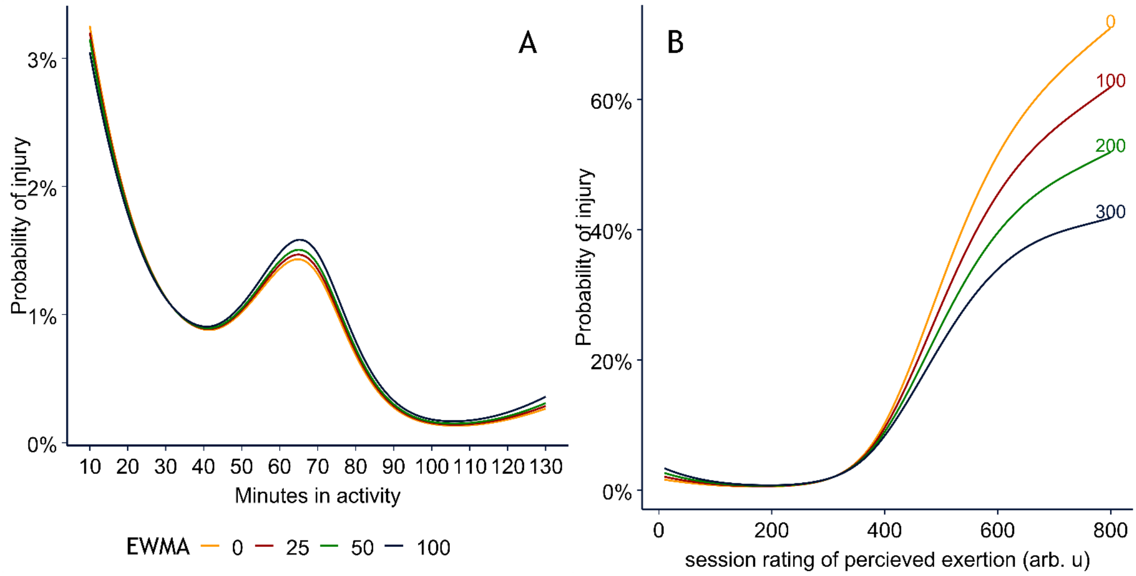

Figure S4. Probability of injury on the current day (Day 0, acute load) predicted by logistic regression models, using EWMA to calculate cumulative chronic load. Shown for each level of training load variables used in (A) Qatar Stars League model (420 329 exposure values, 1 977 injuries) and (B) Norwegian elite U-19 model (4 719 exposure values, 60 injuries). The probability is shown for zero, low, medium and high EWMA levels. Arb. u = arbitrary units, EWMA = Exponentially Weighed Moving Average.

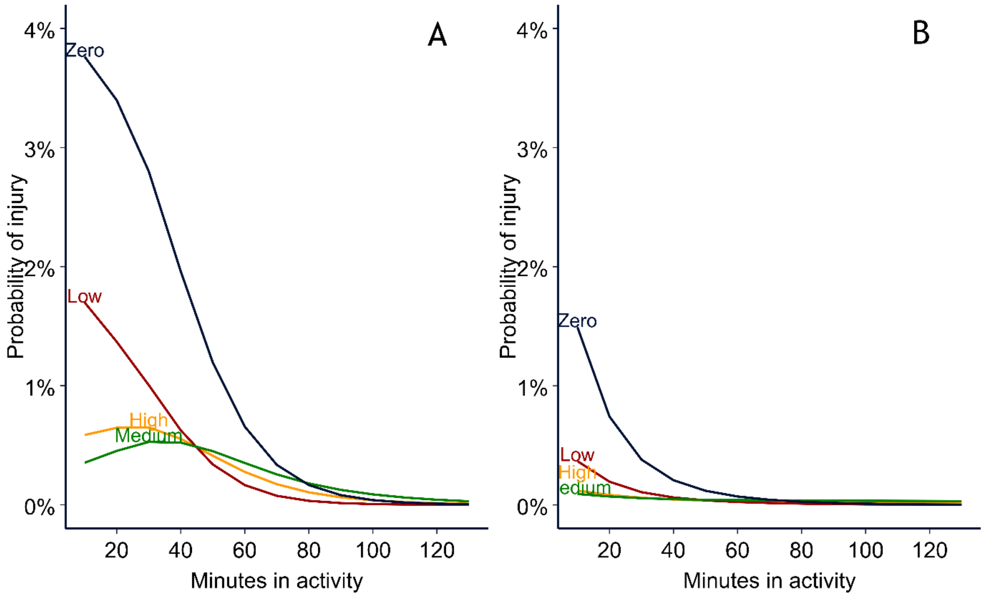

Figure S5. Probability of injury on the current day (Day 0, acute load) for each minute in activity in the Qatar Stars League population (420 329 exposure values), stratified by (A) sudden onset injuries (n = 1 625) and (B) gradual onset injuries (n = 320). The probability is shown for zero, low, medium and high cumulative chronic minutes in activity. The sums chosen were: zero = 27-day sum of 0 minutes, low = 180 minutes, medium = 1 435 minutes, high = 1 900 minutes. The exact profiles used are shown in Table S1. The probabilities were predicted by logistic regression with an interaction term between the acute number of minutes in activity (Day 0) and the cumulative chronic number of days in activity (Day -1 to -27).

References

Barzi, F., & Woodward, M. (2004). Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *American journal of epidemiology*, *160*(1), 34-45.

Moons, K. G., Donders, R. A., Stijnen, T., & Harrell Jr, F. E. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of clinical epidemiology*, *59*(10), 1092-1101.

Peters, S. A., Bots, M. L., den Ruijter, H. M., Palmer, M. K., Grobbee, D. E., Crouse III, J. R., O'Leary, D. H., Evans, G. W., Raichlen, J. S., & Moons, K. G. (2012). Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. *Journal of clinical epidemiology*, *65*(6), 686-695.

Van Buuren, S. (2018a). *Flexible imputation of missing data*. Chapman and Hall/CRC.

Van Buuren, S. (2018b). *Flexible Imputation of Missing Data*. CRC Press. Retrieved 2021-02-12 from https://stefvanbuuren.name/fimd/

Appendix II: Norwegian handball documentation

Norwegian Centre for Research Data Approval (in Norwegian)

Consent form (in Norwegian)

NSD

Norges idrettshøgskole
Att: Christian Thue-Bjørndal
christian.bjorndal@NiH.no

Vår dato: 29.8.2018 Vår ref: 60750/PEG/LR Deres dato: Deres ref:

VURDERING AV BEHANDLING AV SÆRSKILTE KATEGORIER PERSONOPPLYSNINGER I PROSJEKTET: EN LONGITUDINELL STUDIE AV SPILLERUTVIKLING I NORSK HÅNDBALL

NSD - Norsk senter for forskningsdata AS viser til meldeskjema innsendt 11.05.2018. Meldingen gjelder behandling av personopplysninger til forskningsformål.

Etter avtale med den behandlingsansvarlige, Norges idrettshøgskole (heretter NiH), har NSD foretatt en vurdering av om den planlagte behandlingen er i samsvar med personvernlovgivningen.

Resultat av NSDs vurdering:

NSD vurderer at det vil bli behandlet særskilte kategorier personopplysninger om helse frem til 31.12.2018.

NSDs vurdering er at behandlingen vil være i samsvar med personvernlovgivingen, og at lovlig grunnlag for behandlingen er samtykke.

Vår vurdering forutsetter at prosjektansvarlig behandler personopplysninger i tråd med:
- opplysninger gitt i meldeskjema og øvrig dokumentasjon
- dialog med NSD, og vår vurdering (se under)
- NiH sine retningslinjer for datasikkerhet, herunder regler om hvilke tekniske hjelpemidler det er tillatt å bruke

Nærmere begrunnelse for NSDs vurdering:

1. Beskrivelse av den planlagte behandlingen av personopplysninger

Formålet med prosjektet er å undersøke utviklingsforløp som leder til ulike utfall med hensyn til frafall eller fortsatt idrettsdeltagelse blant unge norske håndballspillere i alderen 13-19 år. Studien fokuserer spesielt på sammenhengen mellom hvordan (a) utøverens treningshverdag og treningsbelastning, (b) risiko for idrettsskader og (c) motivasjon og trivsel, utvikler seg gjennom ungdomsårene, samt relasjoner mellom trener og utøver.

Utvalget består av håndballspiller i alderen 16-19 år, samt deres trenere. Det samlede antallet er 250 respondenter. Utvalget rekrutteres via spillerutviklingstiltak i regi av Norges håndballforbund og via klubbesøk der muntlig og skriftlig informasjon blir gitt. Det samles inn data fra spillerne ved hjelp av elektronisk spørreskjema, og senere personlig intervju. For trenerne samles det inn data ved fokusgruppeintervju. Det fremgår av meldeskjema at det vil behandles sensitive opplysninger om helseforhold om spillerne, både i form av fysiologiske opplysninger, samt opplysninger om psykososiale forhold.

Datamaterialet oppbevares på NiH (nærmere om dette under punkt 5).

Prosjektslutt er, ifølge meldeskjema, 31.12.2020. Det fremgår av meldeskjema at du vil lagre datamaterialet med personopplysninger 31.12.2028 for oppfølgingsstudier/ny forskning. Se mer om dette under punkt 6.

2. Personvernprinsipper

NSDs vurdering er at behandlingen følger personvernprinsippene, ved at personopplysninger;

- skal behandles på en lovlig, rettferdig og åpen måte med hensyn til den registrerte (se punkt 3 og 4)
- skal samles inn for spesifikke, uttrykkelig angitte og berettigede formål og der personopplysningene ikke viderebehandles på en måte som er uforenelig med formålet (se punkt 1 og 3)
- vil være adekvate, relevante og begrenset til det som er nødvendig for formålet de behandles for (se punkt 6)
- skal lagres slik måte at det ikke er mulig å identifisere de registrerte lengre enn det som er nødvendig for formålet (se punkt 5 og 6)

3. Lovlig grunnlag for å behandle særskilte kategorier personopplysninger

Særskilte kategorier - Samtykke ((art. 6.1. a), art. 9.2 a), § 10)
Det fremgår av meldeskjema vi har fått tilsendt at det vil bli innhentet samtykke fra de registrerte. NSD vurderer at den planlagte behandlingen av personopplysninger er lovlig fordi:

- det skal innhentes uttrykkelig samtykke fra de registrerte og
- forsker har oppfylt den særskilte plikten til å rådføre seg med en personvernrådgiver

Ifølge prosjektmeldingen skal ungdommer over 16 år samtykke selv til å delta i prosjektet. Hovedregelen når det registreres sensitive opplysninger til forskningsformål om ungdom under 18 år, er at det må innhentes samtykke fra foreldrene. I dette prosjektet vurderer NSD imidlertid at ungdommer over 16 år kan samtykke til deltakelse på selvstendig grunnlag. Dette ut fra en helhetsvurdering av opplysningenes art og omfang. Vi viser til at ungdom i denne alderen har selvbestemmelse på en rekke områder, de kan bl.a. selv velge utdanning, samtykke til helsehjelp, de er over den seksuelle lavalder, og de kan selv melde seg inn/ut av foreninger. Det er personvernombudets vurdering at ungdommene på 16 år og eldre i dette prosjektet har forutsetninger for å forstå hva deltagelse innebærer.

Samtykket innhentes og dokumenteres via skriftlige samtykkeskjemaer.

4. De registrertes rettigheter

NSD vurderer at den registrerte har krav på å benytte seg av følgende rettigheter: Informasjon, innsyn, retting og sletting av personopplysninger, dataportabilitet begrensning.

Samtykket kan trekkes tilbake ved å henvende seg til prosjektleder Christian Bjørndal, evt. ved å henvende seg til Seksjon for personverntjenester ved NSD.

NSD finner at informasjonsskrivet vil gi de registrerte god informasjon om hva behandlingen innebærer og om hvilke rettigheter de har. Vi har imidlertid følgende kommentar:

- Saksbehandlers navn skal fjernes fra informasjonsskrivet. Det er NSD som i dette tilfellet er personvernrådgiver for NiH, ikke saksbehandleren personlig.
- At eventuell videre lagring for forskningsformål (jf. meldeskjema) kun kan skje med deltakernes eksplisitte samtykke

5. Informasjonssikkerhet

I følge meldingen skal det behandles både indirekte og direkte personopplysningene behandles ved hjelp av elektroniske spørreskjemadata, herunder demografiske opplysninger og opplysninger om helseforhold hos håndballspillerne, samt personlige intervju og fokusgruppeintervjuer m/lydopptak. Det skal opprette koblingsnøkkel/navneliste.

Datamaterialet overføres i sin helhet til en ekstern harddisk innelåst på et kontor på NiH. Harddisken er ikke tilknyttet et nettverk. Navneliste/koblingsnøkkel oppbevares på stasjonær PC kontoret til en prosjektmedarbeider, i enn annen avdeling av NiH enn der harddisken oppbevares.

NSD forutsetter at personopplysningene behandles i tråd med personvernforordningens krav og institusjonens retningslinjer for informasjonssikkerhet.

6. Varighet

Ifølge meldeskjema skal personopplysninger behandles frem til 31.12.2020. Ifølge informasjonsskrivet skal opplysninger som kan knyttes til en enkeltperson da slettes/anonymiseres. Ifølge meldeskjema skal imidlertid personopplysninger oppbevares for oppfølgingsforskning frem til 2028. NSD vil, ved prosjektslutt, be om en oppdatering på om fortsatt lagring er aktuelt, og minner om at dette i så fall må komme tydelig frem i informasjonsskrivet for at samtykket skal anses å dekke slik lagring.

Når det gjelder anonymisering, må NiH kunne dokumentere at datamaterialet er anonymisert. Anonymisering innebærer å bearbeide datamaterialet slik at ingen enkeltpersoner kan bli identifisert. Det gjøres ved å: (ta vekk det som ikke passer, presiser gjerne)

- Slette navn, adresse, telefonnummer, epostadresse, IP-adresse og andre nettidentifikatorer
- Slette eller grovkategorisere alder, bosted, arbeidssted, institusjon, diagnose, lokaliseringsdata og andre bakgrunnsopplysninger
- Slette eller sladde lydopptak.

Meld fra om endringer

Dersom behandlingen av personopplysninger endrer seg, kan det være nødvending å melde dette til NSD via Min side. På våre nettsider informerer vi om hvilke endringer som må meldes. Vent på svar før endringen gjennomføres.

Informasjon om behandlingen publiseres på Min side, Meldingsarkivet og nettsider

Alle relevante saksopplysninger og dokumenter er tilgjengelig:

- via Min side for forskere, veiledere og studenter

- via Meldingsarkivet for ansatte med internkontrolloppgaver ved NiH

NSD tar kontakt om status for behandling av personopplysninger
Etter avtale med NiH vil NSD følge opp behandlingen av personopplysninger underveis og ved planlagt avslutning.

Vi sender da en skriftlig henvendelse til prosjektansvarlig og ber om skriftlig svar på status for behandling av personopplysninger.

Se våre nettsider eller ta kontakt ved spørsmål. Vi ønsker lykke til med behandlingen av personopplysninger.

Med vennlig hilsen

Marianne Høgetveit Myhren
seksjonsleder

Pernille Ekornrud Grøndal
rådgiver

Lovhenvisninger
NSDs vurdering er at den planlagte behandlingen av personopplysninger:

- er regulert av personopplysningsloven, jf. § 2.
- oppfyller prinsippene i personvernforordningen om:
  - lovlighet, rettferdighet og åpenhet jf. art. 5.1 a)
  - formålsbegrensning jf. art. 5.1 b)
  - dataminimering jf. art. 5.1 c)
  - lagringsbegrensning jf. art. 5.1 e).
- kan finne sted med hjemmel i personvernforordningen art. 6.1 a), art. 9.2 a)
- gjennomføres på en måte som ivaretar de registrertes rettigheter personvernforordningen **art. 11-22**

NSD legger til grunn at institusjonen også sørger for at behandlingen gjennomføres i samsvar med personvernforordningen:

- art. 5.1 d) og art. 5.1. f) og art. 32 om sikkerhet

Vil du delta i forskningsprosjektet
"Spillerutvikling i norsk håndball"?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å undersøke ulike utviklingsforløp gjennom ungdomshåndballen, frafall og klubbene og idrettsskolene sin rolle i spillerutvikling i norsk håndball. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Vi er spesielt interessert i hvordan du som spiller opplever forholdet mellom trenings- og konkurransehverdagen, motivasjon og trivsel, og håndballrelaterte helseplager. Vi er også interessert i treneres opplevelse av god spillerutvikling. Derfor ønsker vi å rekruttere håndballspillere i alderen 13-19 år som vi kan følge gjennom ungdomshåndballen, og trenere med erfaring fra samme aldersgruppe. Forskningsprosjektet gjennomføres i sin helhet i regi av Norges Idrettshøgskole.

Hvem er ansvarlig for forskningsprosjektet?

Prosjektleder for studien er Christian Thue Bjørndal. Dersom du ønsker å delta eller har spørsmål til studien, ta kontakt med ham på telefon 408 98 766.

Hvorfor får du spørsmål om å delta?

Du får spørsmål om å delta i studien fordi du er mellom 13-19 år og spiller håndball i klubb og/eller går på skole med håndballtilbud.

Hva innebærer det for deg å delta?

Hvis du velger å delta i studien så innebærer det at du vil få tilsendt en link til en nettbasert spørreundersøkelse en gang i året. Det vil ta mellom ca. 15 og 30 minutter. Noen vil også bli forespurt om å svare på en ukentlig eller daglig undersøkelse over en kortere tidsperiode. Det vil ta deg ca. 5-10 minutter. Spørsmålene vil omhandle din opplevelse av trenings- og konkurransehverdagen, motivasjon og trivsel, og håndballrelaterte helseplager. Dine svar fra spørreskjemaet vil bli registrert elektronisk.

I tillegg vil du kunne bli kontaktet med forespørsel om å stille til intervju alene eller sammen med andre spillere. Vi tar lydopptak og notater fra intervjuene. Foreldre kan på forespørsel få tilsendt spørreskjema og intervjuguide i forkant.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Svarene dine vil ikke få noen konsekvenser for håndballspillingen din eller forholdet ditt til treneren e.l.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket. Det er kun prosjektgruppen ved Norges Idrettshøgskole som vil ha tilgang til personopplysninger. Navnet og kontaktopplysningene dine vil vi erstatte med en kode som lagres på egen navneliste adskilt fra øvrige data. Alle data anonymiseres ved publisering og du vil ikke kunne gjenkjennes.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes 31.12.2020, men du kan også bli forespurt om å være med i en videreføring av studien. Etter prosjektslutt anonymiseres alle data og intervjuopptak.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Norges Idrettshøgskole har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Norges Idrettshøgskole ved Christian Thue Bjørndal på e-post christian.bjorndal@nih.no eller på telefon 408 98 766.
- NSD – Norsk senter for forskningsdata AS, på epost (personvernombudet@nsd.no) eller telefon: 55 58 21 17.

Med vennlig hilsen

Christian Thue Bjørndal

Prosjektansvarlig

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet 'Spillerutvikling i norsk håndball', og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i spørreskjema
- ☐ å delta i intervju
- ☐ at mine personopplysninger lagres etter prosjektslutt, til bruk i oppfølgingsstudier

For ungdom og voksne over 15 år. Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. 31.12.2020

(Signert av prosjektdeltaker, dato)

På vegne av ungdommen under 15 år, så samtykker jeg som forelder/verge på at vi har mottatt informasjon om studien, og er villig til å delta

Navn på ungdommen:
Navn på forelder:

(Signert av forelder/verge, dato)

Appendix III: Qatar Stars League documentation

Approval of study protocol

Approval of data sharing agreement between Aspetar and OSTRC

Anti- Doping Lab Qatar Institutional Review Board

Tel: 44132988
Fax: 44132997

IRB MoPH Registration: SCH-ADL-070
MoPH Assurance: MOPH-A-ADL-Q-071

APPROVAL NOTICE

Date	17th Oct, 2017
Lead Principal Investigator	**Karim Chamari, Aspetar**
Co-PI	**Yorck Olaf, Mokhtar Chaabane, Montassar Tabben, Ramadan Daoud, Raouf Nader Rekik, Roald Bahr, Souhail Chebbi**
IRB Application #	E2017000252
Site/s	**Aspetar**
Funding Entity	**Aspetar**
Protocol Title	**Injury and Illness epidemiology in professional soccer players in Qatar: A 10-year longitudinal study**
Submission Type	**Initial Submission**
Review Type	**Expedited**
Approval Period	17th Oct, 2017 – 16th Oct, 2018

The Anti-Doping Lab Qatar Institutional Review Board has reviewed and approved the above referenced protocol. As the Principal Investigator of this research project, you are responsible for:

- Ethical compliance and protection of the rights, safety and welfare of human subjects involved in this research project.
- To follow the policies and procedures as set by ADLQ-IRB in any matters related to the project, following the ADLQ-IRB approval which includes:-
  - Obtaining prior approval of any modifications to the approved protocol including the change of research team members.
  - Reporting deviations and unanticipated events; major deviations within 24 hours.
  - Renewing Ethics annually or every six months if IRB requires it.
  - Submission of progress reports annually
  - Informing the ADLQ-RO of the date of commencement of the research.

ADLQ IRB Chair
Dr. Yorck Olaf Schumacher

مختبر مكافحة المنشطات قطر
Anti Doping Lab Qatar

**For Commencement of Research, Protocol Deviation Reporting, Unanticipated Problem Reporting & Research Progress Annual Report, please contact - Education & Research Office, Anti-Doping Lab Qatar.*

ASPIRE ZONE

AZF Institutional Review Board Approval Notice

MoPH Registration: IRB-AOSM-2020-007
MoPH Assurance: IRB-A-AOSM-2020-0036

Date	**29 September 2022**	
Lead Principal Investigator	**Karim Chamari**	
Co-PI	**Yorck Olaf Schumacher, Mokhtar Chaabane, Montassar Tabben, Ramadan Daoud, Raouf Nader Rekik, Roald Bahr, Souhail Chebbi, Norminda Guanezo, Emna Dachraoui, Zineb Elandaloussi**	
IRB Application #	**E2017000252**	
Sites	**Aspetar**	
Funding Entity	**Aspetar**	
Protocol Title	**Injury and Illness epidemiology in professional soccer players in Qatar: A 10-year longitudinal study**	
IRB Review Type	**Expedited Review**	
Submission Type	**Data sharing request**	
Approval Period	**Valid until the expiry of ethics approval**	**29 September 2023**

AZF IRB APPROVED

The AZF IRB has reviewed and approved the above referenced protocol.
As the Principal Investigator of this research project, you are responsible for:

- Ethical compliance and protection of the rights, safety and welfare of human subjects involved in this research project.
- Adhering to policies and procedures as set by AZF IRB in any matters related to the project, following the AZF IRB approval which includes: -
- Obtaining prior approval of any modifications to the approved protocol including the change of research team members.
- Reporting deviations and unanticipated events; major deviations within 24 hours.
- Informing the IRB Office of the date of commencement of the research.
- LPI may use the content of the approved Informed Consent form in their own organizational letter head, if it deems fit for the nature of the project.
- Research records must be retained for at least 3 years after completion of the research.
- Renewing Ethics annually or every six months if IRB requires it.
- Submission of progress reports annually.
- Failure to renew ethics prior to the expiry date will lead to a cessation notice. Continuing human Subject procedures after the expiry date is a violation.

02/10/2022

Dr. Yorck Olaf Schumacher
(AZF IRB Chair)

Aspire Zone
PO Box 23833
Doha - Qatar
t +974 44476786
f +974 44472045
www.aspirezone.qa

أعضاء
مؤسسة أسباير زون
Members of
Aspire Zone Foundation

أكاديمية أسباير
Aspire Academy
سبيتار مستشفى جراحة العظام والطب الرياضي
Aspetar Orthopaedic and Sports Medicine Hospital
أسباير لوجيستكس
Aspire Logistics

Lena Kristin Bache-Mathiesen // Improving the Methodology of Training Load and Injury Risk Research