2019

Time-to-event analysis for sports injury research part 1: Time-varying exposures

Rasmus Oestergaard Nielsen
Michael Løjbach Bertelsen
Daniel Ramskov
Merete Møller
Adam Hulme

See next page for additional authors

10.1136/bjsports-2018-099408

Originally published as:

Original article available [here](https://ro.ecu.edu.au/ecuworkspost2013/5587).

This Journal Article is posted at Research Online.
Authors
Rasmus Oestergaard Nielsen, Michael Lejbach Bertelsen, Daniel Ramskov, Merete Møller, Adam Hulme, Daniel Theisen, Caroline F. Finch, Lauren Victoria Fortington, Mohammad Ali Mansournia, and Erik Thorlund Parner

This journal article is available at Research Online: https://ro.ecu.edu.au/ecuworkspost2013/5587
Time-to-event analysis for sports injury research part 1: time-varying exposures

Rasmus Oestergaard Nielsen,1 Michael Lejbach Bertelsen,1 Daniel Ramskov,1,2 Merete Møller,3 Adam Helme,4 Daniel Theisen,5 Caroline F Finch,6 Lauren Victoria Fortington,6,7 Mohammad Ali Mansournia,8,9 Erik Thorlund Parner10

ABSTRACT

Background ‘How much change in training load is too much before injury is sustained, among different athletes?’ is a key question in sports medicine and sports science. To address this question the investigator/practitioner must analyse exposure variables that change over time, such as change in training load. Very few studies have included time-varying exposures (eg, training load) and time-varying effect-measure modifiers (eg, previous injury, biomechanics, sleep/stress) when studying sports injury aetiology.

Aim To discuss advanced statistical methods suitable for the complex analysis of time-varying exposures such as changes in training load and injury-related outcomes.

Content Time-varying exposures and time-varying effect-measure modifiers can be used in time-to-event models to investigate sport injury aetiology. We address four key-questions (i) Does time-to-event modelling allow change in training load to be included as a time-varying exposure for sport injury development? (ii) Why is time-to-event analysis superior to other analytical concepts when analysing training-load related data that changes status over time? (iii) How can researchers include change in training load in a time-to-event analysis? and, (iv) Are researchers able to include other time-varying variables into time-to-event analyses? We emphasise that cleaning datasets, setting up the data, performing analyses with time-varying variables and interpreting the results is time-consuming, and requires dedication. It may need you to ask for assistance from methodological peers as the analytical approaches presented this paper require specialist knowledge and well-honed statistical skills.

Conclusion To increase knowledge about the association between changes in training load and injury, we encourage sports injury researchers to collaborate with statisticians and/or methodological epidemiologists to carefully consider applying time-to-event models to prospective sports injury data. This will ensure appropriate interpretation of time-to-event data.

INTRODUCTION

In the past decades, general methodologists of science insisted that it was impossible to measure how health-related exposures and outcomes changed over time.1 Rather, researchers interested in the study of change were encouraged to ‘frame their questions in other ways’.1 Clearly this was a poor advice. Today, one of the overarching goals of sports injury research is to understand why, and when, athletes sustain injury. A current example of this is the ‘too much, too soon’ theory.2

This theory proposes that athletes are at greater risk of injury following a sudden change in training load, a sudden change in the magnitude of the load, a sudden change in the way the load is distributed, or a combination of these changes.3 Clearly, the study of change over time is crucial in the sport injury context. Refining the concept, the sports injury community has become increasingly interested in research questions such as ‘how much change in training load is too much before injury is sustained, among athletes with different characteristics?’3-5 as it is now feasible to measure how individual training loads change over time, using wearable devices with Global Positioning System to facilitate cost-efficient data collection. Across many sports, researchers collect longitudinal data, for example, on training load and injury occurrence, from hundreds of athletes over a full season, a full year, or ideally, longer periods.5 So, how should these data be analysed to study the impact of changes over time on the development of sports injury?

To study change over time, time-varying exposures (eg, change in training load) and time-varying outcomes (eg, change in injury status) are two essential concepts.5 With these concepts in mind, analysing the association between changes in training load and the onset of injury has received careful attention in the scientific literature. In fact, a plethora of original articles5 and consensus reports2 7 were published during the years 2015–2017 alone. In a systematic review8 of 31 articles that examined the training load-injury relationship, X² tests and logistic regression8 9 were identified as the most commonly used analytical approaches, whereas other approaches such as time-to-event models were used in only two articles (6%)10 11 (table 1). These findings have been confirmed in a recent methodological paper as less than 10% of all results in the identified studies were based on time-to-event or multilevel modelling.12

The importance of this finding is that, unless the time to injury is discretised (eg, transferring continuous timescales or variables into units for analyses), it is not possible to include time-varying exposures in traditional logistic regression models or X² tests. Since logistic regression has been the primary choice of many sports injury researchers, this initiates an important debate: How well has the ‘too much change in training load, too soon’ theory been explored in the existing literature? And how reliable are the results that have been reported to date? To facilitate more refined insights into sports injury occurrence,2 we should carefully consider which analytical approach best assesses associations between changes in training load and injury onset.

Digging deeper into the concept of time-varying exposures, sports injury researchers (particularly methodologists, biostatisticians and epidemiologists) can learn from the broad biostatistical subdiscipline called time-to-event modelling. Time-to-event modelling allows analysis of changes in training load and
Table 1  Overview of the statistical methods, as stated by the authors in the statistics section in each manuscript, used to examine the association between training load and injury in 31 original articles included in a systematic review by Drew and Finch.

<table>
<thead>
<tr>
<th>First author (reference number)</th>
<th>Year</th>
<th>Sample size</th>
<th>Statistical method*</th>
<th>Supplementary methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rugby and rugby union</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gabbett (master’s thesis)</td>
<td>2004</td>
<td>79</td>
<td>$X^2$</td>
<td>One-way analysis of variance</td>
</tr>
<tr>
<td>Gabbett</td>
<td>2004</td>
<td>220</td>
<td>$X^2$</td>
<td>One-way analysis of variance</td>
</tr>
<tr>
<td>Gabbett (2005)</td>
<td>2005</td>
<td>153</td>
<td>GLM (OR)</td>
<td></td>
</tr>
<tr>
<td>Brooks</td>
<td>2008</td>
<td>502</td>
<td>Pearson correlation</td>
<td></td>
</tr>
<tr>
<td>Gabbett (2010)</td>
<td>2010</td>
<td>91</td>
<td>Logistic regression</td>
<td>Two-way analysis of variance</td>
</tr>
<tr>
<td>Killen</td>
<td>2010</td>
<td>36</td>
<td>$X^2$</td>
<td></td>
</tr>
<tr>
<td>Gabbett (2011)</td>
<td>2011</td>
<td>79</td>
<td>Pearson correlation</td>
<td></td>
</tr>
<tr>
<td>Gabbett (2012)</td>
<td>2012</td>
<td>34</td>
<td>Time to event (Cox)</td>
<td></td>
</tr>
<tr>
<td>Huilin</td>
<td>2016</td>
<td>53</td>
<td>Logistic regression</td>
<td></td>
</tr>
<tr>
<td>Cross</td>
<td>2016</td>
<td>173</td>
<td>GLM</td>
<td></td>
</tr>
<tr>
<td>Cricket</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dennis</td>
<td>2003</td>
<td>90</td>
<td>2×2 tables (risk ratio)</td>
<td>T-test</td>
</tr>
<tr>
<td>Dennis (2005)</td>
<td>2005</td>
<td>44</td>
<td>2×2 tables (risk ratio)</td>
<td></td>
</tr>
<tr>
<td>Orchard</td>
<td>2009</td>
<td>129</td>
<td>2×2 tables (risk ratio)</td>
<td>T-test</td>
</tr>
<tr>
<td>Huilin (2014)</td>
<td>2014</td>
<td>28</td>
<td>Logistic regression</td>
<td></td>
</tr>
<tr>
<td>Orchard (2015)</td>
<td>2015</td>
<td>235</td>
<td>Logistic regression</td>
<td></td>
</tr>
<tr>
<td>Football</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Love</td>
<td>2006</td>
<td>19</td>
<td>Logistic regression</td>
<td></td>
</tr>
<tr>
<td>Piggot (master’s thesis)</td>
<td>2009</td>
<td>16</td>
<td>Pearson correlation</td>
<td></td>
</tr>
<tr>
<td>Brink</td>
<td>2010</td>
<td>53</td>
<td>Multinomial regression</td>
<td></td>
</tr>
<tr>
<td>Rogalski (2013)</td>
<td>2013</td>
<td>46</td>
<td>Logistic regression</td>
<td>$X^2$</td>
</tr>
<tr>
<td>Colby (2014)</td>
<td>2014</td>
<td>46</td>
<td>Logistic regression</td>
<td>$X^2$</td>
</tr>
<tr>
<td>Ehrmann</td>
<td>2015</td>
<td>19</td>
<td>Unable to assess article</td>
<td></td>
</tr>
<tr>
<td>Other sports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lyman (2001)</td>
<td>2001</td>
<td>398</td>
<td>GLM</td>
<td></td>
</tr>
<tr>
<td>Lyman (2002)</td>
<td>2002</td>
<td>476</td>
<td>GLM</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Wilson</td>
<td>2010</td>
<td>20</td>
<td>$X^2$</td>
<td>Pearson correlation</td>
</tr>
<tr>
<td>Visnes (2013)</td>
<td>2013</td>
<td>141</td>
<td>Logistic regression</td>
<td></td>
</tr>
<tr>
<td>Wheeler</td>
<td>2013</td>
<td>7</td>
<td>Residual maximum likelihood</td>
<td></td>
</tr>
<tr>
<td>Malisoux (2013)</td>
<td>2013</td>
<td>154</td>
<td>Time to event (Cox)</td>
<td></td>
</tr>
<tr>
<td>Bah (2014)</td>
<td>2014</td>
<td>44</td>
<td>Unknown model</td>
<td></td>
</tr>
</tbody>
</table>

Papers in which time-to-event models were used are highlighted in bold. Cox refers to Cox proportional hazards regression. Football includes soccer and Australian football. Other sports: rowing, baseball, basketball, swimming, volleyball and multiple sports.

*Used to examine the association between training load and sports injury. GLM, generalised linear model.

The purpose of this paper is to discuss how changes in training load, which is an example of a time-varying exposure, can be used in time-to-event models to investigate injury aetiology in the sport context. This is part of a strategic editorial commitment by BJSM to advance the quality of methods used in sports medicine research. In this article, part 1 of two articles, we focus on time-varying exposures. Table 2 provides a brief overview of the key questions addressed in this manuscript. Time-varying outcomes, competing risks and subsequent injuries, through a time-to-event lens, are presented in an accompanying paper entitled ‘Time-to-event Analysis for Sports Injury Research Part 2: Time-varying Outcomes’.

### TIME-TO-EVENT AND TIME-VARYING EXPOSURES

One critical feature of prospective, longitudinal sports injury data is that exposures, for example, training patterns, strength, flexibility and behaviour (risk-taking), inevitably change between the time they were measured at baseline and during the follow-up period. Recognising and incorporating such changes into an analysis is required if sports injury aetiology is to be examined. Still, the following questions remain: (1) Does time-to-event modelling allow researchers to include change in training load as a time-varying exposure to sport injury development? (see: Time-varying exposures question 1); (2) Why is time-to-event analysis superior to other analytical concepts? (see: Time-varying exposures...
Table 2  Four common questions in sports injury research and how they can be addressed using time-to-event analysis

| Question 1 | What is a time-varying exposure? | Sports injury researchers are addressing the question ‘How much change in training load is too much before injury is sustained, among athletes with different characteristics?’ Does time-to-event modelling allow them to include change in training load as a time-varying exposure to sport injury development? |
| Question 2 | Why time-to-event modelling? | Why is time-to-event analysis superior to other analytical concepts when analysing training load-related data that changes status over time? |
| Question 3 | Considerations regarding training load-related time-varying exposures | How should sports injury researchers include change in training load in a time-to-event analysis? Key point 3a: In a time-to-event analysis, change in training load can be included as a categorised variable. This enables for examining non-linear dose–response relationships in the association between changes in training load and sports injury. Key point 3b: The acute:chronic workload ratio (ACWR) concept is a concrete example of a categorised time-varying training load. In this, the cut-off points of 0.8 and 1.3 have been suggested. It must be stressed that the most suitable value(s) for the cut-off points and whether changes in training load should be based on absolute and/or relative changes remain areas of uncertainty and discussion. Key point 3c: The sports injury researcher can examine changes in training load using either fixed states and/or transitions between states. Therefore, sports injury researchers are advised to specify up front whether the main objective of the study is to examine injury risk in relation to (1) different (but constant) workload states (eg, low, medium/sweet, high), or (2) the transitions between workload states (eg, moving from high to medium). Many other factors than training load also change status over time (eg, body mass, strength, flexibility). Are sports injury researchers able to include such variables into the time-to-event analysis? Key point 4: Yes. Many variables change over a player’s season or career that can be important to consider in respect to sports injury. These variables can be included as time-varying effect-measure modifiers and/or time-varying confounders. The challenge in this scenario is to have sufficient data to support inclusion of these variables. |
| Question 4 | What about other time-varying exposures? | What are time-varying exposures? How does time-to-event modelling allow you to include change in training load as a time-varying exposure to sport injury development? The concept of change is important as training patterns and athletic participation fluctuate on a monthly, weekly or even a daily basis. As changes in training load vary over time, this variable is a so-called time-varying exposure. Time-to-event modelling is well suited to deal with time-varying exposures and its association with sports injury. In sports science, we assume training spikes (eg, excessive progression in training load) can lead to sports injury. Still, we also acknowledge that athletes have recovery periods (reduced training load). Consequently, the researcher needs to include a time-varying exposure variable that consists of sudden spikes and slight increases and/or reductions in training load over time. Therefore, it is more appropriate to label the time-varying exposure of interest as ‘change in training load’ rather than ‘sudden spikes’ or ‘workload progression.’ The latter two are levels (so-called ‘exposure states’) of exposure, not the exposure variable, which is change in training load. Accordingly, the researcher examines the association between changes in training load and sports injury. If the researcher leaves out ‘change in’ and specifies that she/he examines the association between training load and injury, the time-varying nature of the exposure is not clearly specified. In the literature, there are several examples of studies using time-fixed training load-related exposures. For instance, Walter et al.7 examined the association between weekly mileage and running injury. The women running more than an average mileage per week over a 3-month period exceeding 40 had an increased injury risk of 242% compared with those running below 10 miles/week on average in the preceding 3 months. This enabled Walter et al.7 to identify a subgroup of female runners at increased injury risk. However, it remains open to speculation why the women exceeding an average of 40 miles/week over a 3-month period were more vulnerable to injury. If injury occurs owing to sudden changes in one or more variables, we need to consider these sudden changes. Here, the concept... |

question 2); (3) How can the association between changes in training load and injury be examined using time-to-event models? (see: Time-varying exposures question 3); and (4) How many different types of time-varying exposures can be included in time-to-event modelling (see: Time-varying exposures question 4).
of time-varying exposures is a necessary ingredient.

Key point 1: Time-to-event modelling is well suited to deal with time-varying exposures and its association with sports injury. When using a time-varying training load exposure, the primary exposure of interest must be labelled ‘change in training load’, not ‘training load’. Sudden spikes and reductions in training load are not exposure variables, but exposure levels (known as ‘states’). Consequently, researchers do not examine the association between sudden spikes in training load and injury. They examine the association between changes in training load and sports injury.

Time-varying exposures question 2: why time-to-event modelling?

Regardless of whether changes in training load are defined based on relative changes, for example, the acute:chronic workload ratio (ACWR),27 a modified version of the ACWR,10 28 29 biweekly changes,30 or absolute changes, a high-quality statistical analysis is needed in order to take into account that each athlete can change their status (in statistical terms: transit between states) many times during the course of follow-up. In the present paper, we guide researchers towards time to event when choosing a statistical model. Still, one may speculate: Why is time-to-event analysis superior to other analytical concepts when analysing training load-related data that changes status over time?

In prospective studies, researchers collect data from the same individuals over time. Most likely, sports injury researchers will experience that the training load data from each individual will vary from the time of inclusion (baseline) and during the follow-up. Having such data allows sports injury researchers to model and compare the injury risks across athletes being in different training ‘zones’ or ‘levels’ (statistically: states) over time. For instance, one could examine if injury risk is greater following a sudden increase in training load compared with a slight increase in training load. Here, simple logistic regression and $\chi^2$ test are too restricted to be able to provide answers. Advanced analytical techniques are required. In time-to-event models, the researcher is able to compare injury risk across different changes in training load using what is termed a ‘delayed entry function’. Then, an individual should only be considered at risk in the time period at which the individual is in the given state. Knowing the concept of delayed entry is important as it allows the sports injury researcher to deliver a specific request to the statistician: "We need to analyse changes in each of the athletes’/players’ training load data using delayed entry. Can you help me with that please?"

In addition to acknowledging the importance of dealing with the delayed entry, to model how the injury risk depends on change in training load, time-to-event models allow the possibility of censoring participants as it is likely in prospective studies that some study participants leave the study during follow-up for various reasons. If the researchers do not take into account censoring, they assume that all participants in the study complete follow-up. This is also a very speculative assumption.
Taking into account varying training data and censoring is possible in time-to-event modelling, and it is possible to include time-varying exposures. These strengths make time-to-event modelling vastly more sophisticated than, for example, logistic regression analysis in which the analysis is limited to the inclusion of time-fixed variables and is unable to use delayed entry. In addition, the censoring of participants in logistic regression requires either adjustment for length of follow-up, which leads to biased estimates, or additional statistical programming because the codes to run the analyses are not included in most readily available software. Consequently, time-to-event modelling should be considered as a preferred analytical strategy in sports injury research when examining the association between changes in training load and injury occurrence—at least when the outcome is dichotomised or categorised.

**Key point 2:** In contrast to logistic regression and X² test, time-to-event modelling allows the researcher to take into account censoring and compare injury risk across time-varying exposures by using delayed entry functions.

It is important to recognise that alternative methods to handle time-varying exposures exist. Other modelling strategies (eg, generalised estimating equations, random effects models or multilevel regression approaches) can be used as they, like time to event, take into account the repeated measures by clustering observations over time within individuals. Consequently, generalised estimating equations, random effects models and/or multilevel regression approaches also provide researchers with opportunities to analyse data based on repeated measurements and with within-subject correlation. In those analyses, the outcome can be categorical or quantitative/continuous. A description of these methods can be found elsewhere, for example, in Smith and Walls study. To the best of our knowledge, readily available software does not allow researchers to deal as easy with the concept of exposure variables, like changes in training load, which changes status over time when using generalised estimating equations, random effects models and/or multilevel regression approaches. Therefore, time-to-event models are a more feasible approach for most sports injury researchers.

**Time-varying exposures question 3: how to include changes in training load in a time-to-event analysis?**

When the researcher has chosen to include changes in training load as a time-varying exposure variable and use time to event as analytical approach, there several practical options remain:

- **Continuous versus categorised exposure:** A priori one could assume that there is a linear relationship between changes in training load and injury risk. However, it is reasonable to question whether or not a linear dose–response relationship exists in the association between changes in training load and injury. If the relationship is assumed to be non-linear, the next step becomes to categorise time-varying exposure based on a set of a priori defined cut-offs (perhaps three or four ‘states’ of change). Although spline regression and fractional polynomials have shown promising results when handling continuous training load exposures, here, we limit ourselves to categories with cut-offs, which allows for the examination of exponentialised relationships and/or U-shaped patterns to describe associations between changes in training load and injury.

**Key point 3a:** In a time-to-event analysis, change in training load can be included as a categorised variable. This enables examining non-linear dose–response relationships in the association between changes in training load and sports injury.

- **Defining cut-offs:** Choosing cut-offs will categorise the exposure variable, in this case changes in training load, into certain exposure states. Note: the term ‘exposure states’ and not ‘exposure groups’. In exposure groups, each individual is assigned to a certain time-fixed exposure group without the possibility to transit between groups. Importantly, ’state’ indicates that each individual has the opportunity to switch/transit between exposure states during follow-up.

To define cut-offs, we encourage researchers to use appropriate, up-to-date scientific/biological rationale as the basis for their choices instead of allowing the statistical software program to choose the cut-offs based on data-driven knots. As these data-driven knots are produced by a software program, the approach leaves little room for hypothesis-driven research. Taking the concept of the ACWR as a concrete example of time-varying training load, we know that cut-offs of 0.8 and 1.3 have been suggested as being relevant and in other articles, authors have suggested using load progressions of 10%, 20%, 30% or 60% as indicating critical change in load. Although these different ways of categorising changes in training load are appealing to use in future studies, it remains uncertain which cut-offs are appropriate. This leaves the researcher with many possibilities for choosing the cut-offs they believe, based on subject matter knowledge and studies from the literature, are most appealing.

**Key point 3b:** The ACWR concept is a concrete example of time-varying training load. Here, the cut-offs of 0.8 and 1.3 have been suggested as important cut-offs. Still, it must be stressed that it is uncertain as to which cut-offs are suitable and if changes in training load should be based on absolute and/or relative changes.

- **Choosing between states and transitions:** Analysts should decide if they want to analyse changes based on exposure states themselves or transitions between exposure states, as these are two different concepts and both are valid. If change in the ACWR is used as the primary exposure of interest, and the corresponding cut-offs are set at 0.8 and 1.3, then the researcher considers three exposure states: below 0.8, between 0.8 and 1.3, and above 1.3 (figure 1). This is appropriate when the researcher aims to examine whether a certain state (eg, the ‘sweet spot’ between 0.8 and 1.3) is associated with greater/lower injury risk compared with being in another state (below 0.8 or above 1.3). On the other hand, asking whether transitions between two states convey changes in injury risk is a question of interest that the previous approach does not answer. Specifically, is it more injurious to switch from the sweet spot (say, ‘between 0.8 and 1.3’) to an ACWR of >1.3, rather than remaining constantly in the 0.8–1.3 sweet spot. With three exposure states (below 0.8, between 0.8 and 1.3, and above 1.3) there are nine different options for multistate transitions (MST):  
  - MST1: Below 0.8 and remaining below 0.8.  
  - MST2: Below 0.8 to between 0.8 and 1.3.  
  - MST3: Below 0.8 to above 1.3.  
  - MST4: Between 0.8 and 1.3 to below 0.8.  
  - MST5: Between 0.8 and 1.3 and remaining between 0.8 and 1.3.  
  - MST6: Between 0.8 and 1.3 to above 1.3.
of players with different levels of scapular control. In addition to including changes in training load as a time-varying exposure, scapular control was included as a time-varying effect-measure modifier—not just as a baseline value. This was under the assumption that the neuromuscular function supporting scapular control among handball players will not be constant across one or more seasons—and thus the duration of the study. The biological rationale was that scapular control can change during a season because of either (1) purposeful rehabilitation, (2) muscle inhibition/imbalance over the course of a season because of, for example, muscle fatigue, or (3) frank injury to the rotator cuff (shoulder tendinopathy) or another body part that limits the player’s ability to train.

In the handball study above, the data set-up included a time-varying training load exposure and a time-varying non-training-related variable (scapular control). This is an example of the concept of effect-measure modification. By using the concept of effect-measure modification, the researcher is able to examine how much change in training load is too much among athletes with different characteristics.

When a researcher is interested in examining the combined effect of two exposures that both are related to change in training load, say change in running distance and change in running intensity, this is also possible via time-to-event modelling using interaction, not effect-measure modification or confounding, between the two time-varying training load-related variables. Such an analytical approach is suitable if a researcher wants to examine questions such as: Is it more injurious to progress in running distance and in running intensity at the same time compared with progressing in running distance while running at the same intensity? For an example of such a set-up, we guide the reader to look deeper into table 2 in the related article entitled ‘Time-to-Event Analysis for Sports Injury Research Part 2: Time-Varying Outcomes’.

Key point 4: Many risk factors for sports injury, for example, training load, body mass, strength and flexibility, can be included in the same time-to-event analysis as time-varying training load exposures (and the interaction between them), time-varying effect-measure modifiers and/ or time-varying confounders.

Other research questions

The goal of many sport injury researchers has been to predict injury risk in subgroups of athletes who present with a certain characteristic (or putative biologic exposure). To do this, researchers need to carefully consider the research question ‘How much change in training load is too much before injury is sustained, among athletes with different characteristics?’ and the concepts of time-varying exposures and time-varying effect-measure modification. If the sport injury researcher has the different, but equally important, research question ‘What is the average causal effect of body mass on sports injury occurrence?’ then other analytical approaches need consideration, namely: time-varying confounding.

Time-varying confounding

Over the past decade, techniques underpinning causal inference have emerged. Here, the goal is not to investigate stratum-specific differences but to estimate the causal effect of an exposure on sports injury after adjustment for confounding. Confusing the effect of interest with non-causal associations which results from a common cause of both the exposure and outcome. All study designs, including randomised controlled trials, are subject to random confounding and compliance problems. A minimally sufficient set of confounders (ie, a set of confounders sufficient for confounding elimination of which no subset is sufficient) must be identified via causal diagrams. Then, one approach could be to adjust for the confounders using standard regression modelling. In practice, many exposures of interest are time varying, and the values of potential confounders may change over time leading to time-varying confounding. However, standard regression methods for analysis of longitudinal data such as time-dependent Cox regression do not appropriately adjust for time-varying confounding, and causal methods including inverse probability of treatment weighting, the parametric G-formula and G-estimation, or collectively G-methods should be used instead.

Recent developments within statistics have opened the use of G-methods in time-to-event analysis using proportion-based measures of association. Here, one needs to address assumptions regarding right censoring, since special techniques to estimate average causal effects are required. The pseudo-observation method has proved valuable for this purpose when applying direct
standardisation (G-formula) or inverse probability weights (based on propensity scores). Conclusively, sports injury researchers should be aware that estimation of average causal effects is possible when using Cox regression and a generalised linear model (pseudo-observations). Such analyses are complicated and often require collaboration with a statistician.

Time-varying outcomes

In sports injury research, the concepts of time-varying exposures and outcomes appear to have been rarely used in combination with time-to-event models. Although the reasons for this are unknown, it could be due to either a lack of awareness among sport injury researchers about the potential utility of time-to-event models, or alternatively, the perceived difficulty regarding their use in practice. Another reason could certainly be the limited sample size and event per variable, which are related to the amount of sports injuries available in the data set. This issue will be addressed in the follow-up article entitled ‘Time-to-event Analysis for Sports Injury Research Part 2: Time-varying Outcomes’.

CONCLUSION

Careful attention on how to analyse the time-varying relationship between changes in training load and changes in injury status is needed to address the research question ‘How much change in training load is too much, among athletes with different characteristics, before injury is sustained?’ Time-to-event models are suitable for analysing this highly dynamic relationship as they take into account censoring and the within-individual correlation of follow-up data. Naturally, many factors other than training load change status over time. These include, but are not limited to, body mass, equipment usage, sleep and strength. To take into account that player/athlete characteristics change over time, analytical concepts such as time-varying effect-measure modifiers and/or time-varying confounders are important.

Research into sports injuries is undergoing a transformation with increased attention to stronger analytical methods. As these new insights have potential value for sports injury researchers, there is a need to revisit and further elaborate on these analytical concepts. The analytical approaches presented in this paper require specialist knowledge and well-honed statistical skills to master. Cleaning data sets, setting up the data, performing the analyses and interpreting the results are a time-consuming process which requires dedication and, most likely, assistance from methodological peers. To increase knowledge about the association between changes in training load and injury, sports injury researchers are encouraged to collaborate with statisticians and/or methodological epidemiologists to carefully consider applying time-to-event models to their prospective sports injury data and ensure appropriate interpretations of time-to-event data.

**Contributor**

RON drafted the manuscript and all authors reviewed it for important intellectual content. MAM and EP drafted the section about causal inference.

**Funding**

The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests**

None declared.

**Patient consent**

Obtained.

**Ethics approval**

Local ethics committee in the central Denmark region.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

---

**OPEN ACCESS**

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on others to distribute, remix, adapt, build upon this work in accordance with the Creative Commons Attribution license, provided the original work is properly credited. For details of the license, visit: http://creativecommons.org/licenses/by-nc/4.0/

---

**REFERENCES**

Education reviews


37 Huilme A, Finch CF. From mono-causality to systems thinking: a complementary and alternative conceptual approach for better understanding the development and prevention of sports injury. Inj Epidemiol 2015;2:31.


