Training Load and Its Role in Injury Prevention, Part 2: Conceptual and Methodologic Pitfalls

Franco M. Impellizzeri, PhD*; Alan McCall, PhD†; Patrick Ward, PhD‡; Luke Bornn, PhD§; Aaron J. Coutts, PhD*

*Faculty of Health, Human Performance Research Centre and School of Sport, Exercise and Rehabilitation, University of Technology Sydney, Australia; †Arsenal Football Club, London, United Kingdom; ‡Seattle Seahawks, WA; §Sacramento Kings, CA

In part 2 of this clinical commentary, we highlight the conceptual and methodologic pitfalls evident in current training-load–injury research. These limitations make these studies unsuitable for determining how to use new metrics such as acute workload, chronic workload, and their ratio for reducing injury risk. The main overarching concerns are the lack of a conceptual framework and reference models that do not allow for appropriate interpretation of the results to define a causal structure. The lack of any conceptual framework also gives investigators too many degrees of freedom, which can dramatically increase the risk of false discoveries and confirmation bias by forcing the interpretation of results toward common beliefs and accepted training principles. Specifically, we underline methodologic concerns relating to (1) measure of exposures, (2) pitfalls of using ratios, (3) training-load measures, (4) time windows, (5) discretization and reference category, (6) injury definitions, (7) unclear analyses, (8) sample size and generalizability, (9) missing data, and (10) standards and quality of reporting. Given the pitfalls of previous studies, we need to return to our practices before this research influx began, when practitioners relied on traditional training principles (eg, overload progression) and adjusted training loads based on athletes’ responses.1,2 One reason isolated TL measures cannot be used to prevent injury is the lack of investigation surrounding the causal relationships between measures of TL exposure and injury (see part 1). Also, conceptual and methodologic concerns suggest that prior findings may have been influenced by HARKing (hypothesizing after the results were known), P hacking (running several analyses to find significant results), selective reporting, and confirmation bias. These problems can increase the chance of false discoveries and unsupported claims.

Here we present a series of conceptual and methodologic topics that should be considered when interpreting the results of any studies before practical applications are extracted and given too much credit.

LACK OF A REFERENCE CONCEPTUAL FRAMEWORK AND MODEL: SHARK ATTACKS

Although researchers described their aim as investigating (occasionally predicting) the association between TL and injury,3–7 their conclusions and recommendations were commonly reported as if causality had been assessed. Such erroneous interpretations are understandable. In part 1, we used the example of the popular spurious correlation between shark attacks and the amount of ice cream sold. In this example, it is easy to see how association wrongly interpreted as causality can produce confusion and ultimately the wrong recommendation, as it would mean that if we sell less ice cream, fewer shark attacks will occur. The same has happened with TL and injury studies—that is, if we modify the TL, the assumption was that we could reduce or increase injuries, which has never actually been tested. The exclusion of a causal association is the consequence of our previous knowledge of and expertise about that phenomenon. With potential predictors or prognostic factors that are conceptually or empirically linked to injury occurrence, the absence of a framework leads to the belief that any measure of TL reflects some underlying mechanism of injury. Indeed, injuries occur while training. The conclusion that an association between shark attacks and ice cream is unreasonable is based on our knowledge, and determining the role and suitability of a
The lack of any conceptual framework has another unfortunate consequence: it gives researchers too many degrees of freedom, which, when combined with suboptimal analyses and data handling, can dramatically increase the risk of false discoveries and confirmation bias to support common training practices, such as the overload progression principle.

**Measure of Exposures**

Researchers are free to select whatever they want as measures of exposure in their statistical model. For example, other than the acute and chronic TL, they can use the ACWR, week-to-week variation, ratio between low or high chronic TL, and different categories. Calculating associations from multiple combinations of TL measures inflates the risk of type I errors. This was acknowledged by 1 set of authors who calculated 350 hazard ratios using varying time windows (7–28 days) and TL measures (rolling and exponentially weighted averages, ACWR training monotony, and strain).

**Pitfalls of Using Ratios**

As mentioned previously, the most common measure of TL is the ACWR. However, the physiological rationale for choosing this ratio is unknown, and it is not supported by the Banister model. Furthermore, the ratio is well known to be problematic. Lolli et al observed that the ACWR failed to normalize the numerator for the denominator even after controlling for the mathematical coupling. The ratio is indeed used to control for a denominator variable (chronic workload) that is assumed to influence the numerator (acute workload), which is the variable we consider important. This failure adds unnecessary “noise,” increases the risk of artifact, and makes the results difficult to interpret, as the practitioner cannot determine whether the acute or chronic TL is driving the ACWR. Despite this fundamental flaw, this problem has been completely ignored, and few researchers focused on the coupling problem raised by Lolli et al. Unfortunately, these investigators failed in at least 2 aspects: first, they did not use a within-participants analysis, as should be done for ACWR because this metric is calculated at the within-participants level over a time series, and second, they did not understand that the ratio is the problem. Indeed,
Lolli et al\textsuperscript{12} demonstrated that even when uncoupled, the ratio does not normalize.

It was also shown that when the training schedule was not taken into account, false relationships between ACWR and injury were generated. For example, Born et al\textsuperscript{16} created a 1000-season simulation of TL and injury using data from 2 elite sports (Serie A soccer and National Football League football). The simulation was constructed so that the ACWR was to have no relationship with injury (ie, injury was simulated to simply be a function of the training demands on a given day). When the simulation was statistically evaluated using “common approaches,” the ACWR outside of the proposed “sweet spot” (0.8–1.3)\textsuperscript{7} had a significant relationship with injury (6.8% increase in injury in Serie A soccer and 10.5% increase in injury in National Football League football)—similar to what has been reported in previous literature.\textsuperscript{3,6} Interestingly, this relationship existed despite the simulation’s being designed to have no relationship at all. Once the training day was considered in the model, representing a function of the underlying training schedule, the relationship between the ACWR and injury no longer existed. Such results suggest that relationships with the ACWR in earlier studies might be nothing more than statistical artifacts. Specifically, the statement “ACWR predicts injuries” may be better stated as “ACWR predicts the upcoming TL.”

The idea that new metrics such as ACWR are strongly and consistently associated with injury is not actually supported by evidence. Even in the first 2 studies, on which the famous U-shape model presenting the sweet spot was based, Hulin et al\textsuperscript{1} found an unexpected relationship with injury (6.8% increase in injury in Serie A soccer and 10.5% increase in injury in National Football League football)—similar to what has been reported in previous literature.\textsuperscript{3,6} The conceptual and methodologic inadequacy of the ACWR (ie, to “normalize” the acute by the chronic load) has even been indirectly acknowledged by the proponents of the ACWR and other authors\textsuperscript{3,4,26,47} who arbitrarily removed the training weeks in which the chronic load was 1 or 2 standard deviations below the mean from the ACWR analysis. This exclusion was considered necessary to limit “spikes” in the ACWR, as small absolute increases in acute TL at low chronic TL values would result in very high ACWR values. Because a ratio is used when one believes the numerator (ie, acute TL) is the important factor but wants to control for differences in the denominator (ie, chronic TL),\textsuperscript{26} manipulating the denominator by eliminating data (ie, low chronic TL values) should not be necessary. Furthermore, removing lower TL values from the chronic TL measure to limit spikes creates an unrealistic scenario that does not represent the actual TL exposure.

Finally, in a recent study,\textsuperscript{48} investigators showed that the ACWR (ie, the ratio) actually generated statistical artifacts. The concerns raised in this section should warn researchers and practitioners to avoid adopting into their practice methods that have not been properly scrutinized by independent investigators. Additionally, the methodologic concerns we have raised can be used to avoid repeating similar mistakes in the future.

**Training-Load Measures**

Training load can be assessed using various methods and devices. Again, without a conceptual framework, researchers can use whatever is available. Investigators have used measures of internal TL such as the session rating of perceived exertion and measures of external TL such as GPS and inertial sensors, while ignoring (or not considering) that each of these measure different TL constructs.\textsuperscript{39} Furthermore, different TL indicators have also shown different relationships with injury risk, as demonstrated, for example, by Bowen et al\textsuperscript{5,23} or Jaspers et al.\textsuperscript{50} One obvious limitation of GPS is that it cannot be used to quantify any indoor training (eg, gym training). Why most authors who used GPS failed to quantify or acknowledge this point is unknown. We contend that it is highly unlikely that the teams studied did not complete any gym sessions, as preventive exercise programs are fundamental training activities. Because these programs may reduce the risk of injuries,\textsuperscript{51} this should be considered, at minimum, a potential effect-measure modifier. Conversely, such programs may be thought of as confounders because activity completed in the gym can affect the previous or next field training. Although we acknowledge that not all contextual factors can be measured, gym training is a part of the TL and not a contextual factor. Finally, much of the literature is vague as to how to account for nontraining days (eg, days off) in calculating the ACWR. Lack of clarity in reporting makes it difficult to understand the influence of nontraining days on the chronic load, as these days would technically be listed as a TL of 0.\textsuperscript{52}

During soccer, the match activities of the highest-level professional teams are commonly measured using semi-automated camera systems. Combining data coming from different player-tracking systems is then necessary. Indeed, in a recent study of Premier League players, Bowen et al\textsuperscript{5} used both GPS and camera-system data to represent the same load indicator, referring to a study by Buchheit et al\textsuperscript{15} to support the interchangeability. However, Buchheit et al\textsuperscript{13} not only used a different GPS device but they also found a moderate systematic error after applying necessary customized correction algorithms (which reduced systematic bias). An additional concern is that some matches in team sports, in which GPS is frequently used during competition, are also played indoors, where GPS cannot be used. For instance, in AFL, some matches are played indoors, and the activity profile of certain official matches is measured using a radio-frequency identification–based, active real-time location system and not GPS (which is used in the other matches). However, none of the authors who examined the AFL mentioned whether the analyzed data were combined from the 2 systems or the radio-frequency identification measures were excluded. Collectively, these examples indicate that data from different player-tracking systems must be analyzed with care, as
Carey et al. designed a study aimed at defining the best time to measure the chronic load. They investigated the relationship between training load and injury. This approach is exploratory and trial-and-error, in which various combinations of time windows were explored without providing any justification. They appear to have applied an unreasonable and a form of overfitting.

Method 1 commonly refers to injuries occurring in the current week, and method 4 to injuries occurring in subsequent week. The chronic load (eg, preceding 4-week average) is usually calculated including and starting from the weeks of acute load. Abbreviations: AWCR, acute:chronic workload ratio; AU, arbitrary units.

Researchers have used a plethora of time windows without providing any justification. They appear to have applied an exploratory trial-and-error approach, in which various combinations may be considered “just to see what happens.” Investigators have used 1 to 8 weeks as the indicator of chronic (or cumulative) TL and 1 to 2 weeks for acute TL and added ACWRs calculated with as many time windows. Carey et al. designed a study aimed at defining the best time windows. They explored 336 combinations and determined the results differed depending on the windows (and the TL measure). Indeed, for moderate-speed running, the “best” window was 3 days for acute load and 21 days for chronic load, whereas for total distance, the “best” windows were 6 and 24 days, respectively. Results of these studies are difficult to reconcile to a logical framework, especially in the absence of a reference conceptual model. However, this adaptation of the time windows to the available data is methodologically unreasonable and a form of overfitting.

Another concern is the lack of consistency in the methods used to calculate acute weekly TL. Figure 2 shows 4 ways to calculate the acute load using the session rating of perceived exertion as the TL measure. We have presented these methods using Monday-to-Sunday blocks, which is common in many studies. Each method clearly produces different acute load figures (from 1340 to 3450 arbitrary units), which in turn affect other derived measures of changes created by the authors, for reasons not reported, considered the TL completed in the days immediately before the injury as not affecting the injury risk. Also of concern is that the model combined athletes who experienced an injury immediately after the weekly load used for the calculations and athletes who might have been injured as much as 7 days later. This frequent use of time lags is worrying not only for the lack of biological and plausible explanations but also because, if those studies are correct, from a practical point of view, this would mean that no matter what the athlete does after that week, the injury cannot be prevented. Furthermore, when no time lag has been specified, it is not often possible to know how or whether the TL until the moment of the injury was entered in the analysis (method 2). Method 1 (usually referred to as “current-week injuries”) is influenced by when the injury occurred but has been used in only a few studies. Again, some of these methods are confusing, illogical, and often not properly explained or justified. We recommend that, when practitioners read these articles, they ask themselves whether the researchers’ choices were reasonable and supported by any evidence.

**Discretization and Reference Category**

Most authors have categorized continuous measures of exposure and ratios. The limitations of discretization have been reported in the literature. Nevertheless, investigators who used this approach commonly did not explain the reasons for creating those specific categories, which is important considering their influence on the results. No clear rationale describes why researchers have used between 2 (dichotimization) and 7 categories of ACWR and up to 11 bins. Also unknown is why authors have discretized the data using group and not individual values. Such an approach may cause some players to be underrepresented in certain categories and overrepresented in others—and this affects the interpretability of the results. Authors have analyzed the data using different methods of categorization in the same study (eg, z-score or quartile). Furthermore, the creation of discrete variables may lead to additional problems. For example, Carey et al. showed that 3 discretization approaches producing different categories caused 16 to 21 false discovery rates in 100 simulations (assuming no relation with injury risk). At least 1 of the 3 discrete methods showed a false discovery 42 times out of 100. Another major concern highlighted by Carey et al. was that freedom of choice in selecting a reference category almost doubled the risk of a false discovery. In most, if not all, TL-injury studies, neither the categories nor the references have been clearly justified. The main justification for such approaches appears to arise from authors who simply copied the methods and statistical approaches of previous researchers.

Furthermore, the categories and their meaning should be considered in the interpretation, and we are afraid that most did not consider the implications of the selected reference categories. For instance, Fanchini et al. in a study of football players, demonstrated an increased injury risk at a “high” ACWR (>1.26). Specifically, the injury risk was higher for an ACWR >1.26 versus an ACWR <0.78. Thus, the risk was higher in those with an increased load compared with a decreased load. When athletes with an ACWR >1.26 were compared with those who increased their acute load from 2% to 26% (ACWR = 1.02–1.26), the risk was not higher. In other words, if we want to apply these findings assuming a causal relationship that doesn’t exist, we should avoid spikes only if the players’ TL decreased (ACWR < 1) during the preceding...
week. But once the ACWR is >1.02, even doubling or tripling TL would not increase the injury risk. This finding is exactly the same as in the first study using the ACWR: Hulin et al\textsuperscript{3} showed that at ACWR >1.5 and ACWR >2.0, the injury rate was higher than with an ACWR <0.99 but not different than with an ACWR between 1 and 1.49.

**Injury Definitions**

Various classification systems are available, and the International Olympic Committee\textsuperscript{75} has recently presented a consensus document on standardizing injury data collection. However, the possibility that a researcher will arbitrarily select an injury definition without providing an appropriate rationale to justify the choice (again because of the lack of a reference framework and etiologic model) is another source of degrees of freedom that can change both the study results and the interpretation.\textsuperscript{76} Injuries have been defined and used in many ways, including contact, combined contact and noncontact, match loss, both training and match loss, complaints requiring medical attention, complaints requiring training session modification, and combined upper and lower body injuries, often with unspecified severities or days lost. In addition, some investigations involved self-reported injuries, whereas others relied on the medical staff to record injuries. The specific reasons or mechanisms for selecting a particular definition that clearly determines the number of predicted events have never been presented. Some researchers used definitions based on complaints, whose nature may underlie different, if any, TL associations with specific injuries and specific tissues, as needed or it would at least be advisable to examine associations with specific injuries and specific tissues, as was done in 1 study.\textsuperscript{55} Instead, the use of broad injury categories as outcomes assumes common qualitative and quantitative links between all these injuries and TL measures, which is not reasonable.

**Unclear Analyses**

The use of suboptimal statistical analyses in TL-injury studies has been reported in the literature.\textsuperscript{78,79} However, some authors appear to have intermingled various methods for calculating injury risk, relative risk, and likelihood of risk, which makes interpretation very difficult for the reader. For example, in their seminal papers, Hulin et al\textsuperscript{3,4,61} calculated the injury risk as “the number of injuries sustained relative to the number of exposures to each workload classification,”\textsuperscript{48(233)} They reported this as both injury risk (in tables) and likelihood of risk (in figures). Yet this is not an injury risk but rather an injury rate.\textsuperscript{80} They also calculated the relative risks using these rates, which means they actually calculated rate ratios. The researchers provided values for what they labeled injury risk in tables but highlighted significance levels using \(P\) values determined from odds ratios (with logistic regression as a post hoc test). This method is inappropriate because the \(P\) value referred to the odds ratio and not the injury risk. These authors recently replicated the same errors.\textsuperscript{53}

This is not only confusing but a questionable reporting of statistics.\textsuperscript{3,4,61} Unfortunately, the methods from these studies were used as reference for later studies, and the inaccuracies were replicated by others.\textsuperscript{5,6,7,23,58,81,82}

Further, although confidence intervals have often been reported, uncertainty has rarely been discussed or considered.\textsuperscript{87} The investigators\textsuperscript{88,84} who found extremely large confidence intervals at the extremes of the TL-injury relationship (ie, outside the “sweet spot”) suggested that few injuries occurred in these ranges (also indicating sparse data bias) and the results were compatible with any kind of association.

Finally, we should always consider that a common characteristic of these studies is that the analysis was performed on data from athletes who sustained repeated injuries and, therefore, injury recurrence should have been taken into account. Although some researchers attempted to use statistical methods to handle correlated outcomes within individuals (eg, mixed models and generalized estimating equations), a consistent reporting strategy has yet to be adopted. Inconsistent reporting of such complex models makes their interpretation and comparison of outcomes across studies challenging for the practitioner. Recurrence is an important factor when analyzing data in injury research; simply ignoring it is not the best approach.\textsuperscript{83} Also, most exposures, outcomes, confounders, and moderators are time-varying variables. Nielsen et al\textsuperscript{79,84} provided gentle introductions to this topic, and interested readers are advised to refer to these articles.

**Sample Size and Generalizability**

Because in most of the TL-injury research, authors examined data from single teams in a myriad of sports, the studies have low generalizability and should probably be viewed as case studies. Accordingly, the findings cannot be confidently extended to other teams or populations. Most of the investigations also lacked adequate sample sizes,\textsuperscript{85} which was amplified when the samples were split into subcategories (eg, a high ACWR combined with low or high chronic load).\textsuperscript{4,5,23,46} Adequate sample size should not be calculated using only the events-per-variable approach because this method has limitations,\textsuperscript{3,86,87} but authors often do not report any justification for the chosen sample size. Even worse, some researchers reported the post hoc power analysis, which is well known to be inappropriate.\textsuperscript{85} Moreover, no authors reported the number of injuries for each category, and we are concerned that some categories had as few as 1 or 2 injuries. In addition, the number of athletes who sustained repeated injuries was rarely supplied.

**Missing Data**

The literature contains extensive explanations\textsuperscript{89,90} of how to handle missing data, but these have been ignored by many TL-injury investigators. To be imputed, data should be demonstrated to be (or reasonably assumed to be) missing completely...
at random or at least at random. Today, single imputation is an infrequent practice, and when possible, multiple imputations are suggested as an alternative for completing a case analysis.91 In sport science, simple imputation is quite common, with mean group values typically used as the imputation. Unfortunately, this approach has many limitations, and other solutions (eg, substitution, hot and cold deck imputation, regression) should be considered.89–93 However, more concerning extreme imputation strategies, such as estimating the data of an entire season using the data of subsequent seasons,3 have been used in some TL-injury work, which clearly severely biases the results. To improve on these practices, we suggest that researchers report on the proportion of missing data, assess whether any missing values were missed at random, describe how data imputations were made, and provide a justification of the method. A sensitivity analysis to show the effect of imputations would also be desirable and good practice.

### Standard and Quality of Reporting

Several international guidelines are available for standardizing reporting in scientific studies (www.equator-network.org). These guidelines have been developed to promote standardization, transparency, and high quality of reporting. Such information is necessary to provide readers and other researchers pertinent information about the rationale, design, methods, analysis, and interpretation. These guidelines also allow for accurate evaluation of the risk of bias. Unfortunately, as described earlier, the studies on TL and injuries are heterogeneous in their design and methods. Reporting of methods is also unclear in many studies, which makes interpretation and comparison of findings difficult. For example, rigorously developed and well-established reporting guidelines94,95 recommend presenting both the absolute and relative risk because the relative risk alone can be misleading and influence interpretation and subsequent decisions.96 Relative risk can overestimate the effect of the association between exposure and outcome. For this reason, both relative and absolute risk (eg, risk differences) must be supplied.96,97 Unfortunately, with few exceptions,98 the authors of TL-injury studies commonly only provide the relative risk. Based on the absolute risk differences reported by Colby et al,98 even when trying to contextualize these numbers, the absolute changes in the injury risk appear to be very small or negligible. Other than designing adequate studies, it is important to clearly report each study’s methods. Given the increasing interest and research in injury prediction, the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis99 is a valuable tool whose implementation would improve the reporting so that the risk of bias can be assessed and the methods replicated.

### CONCLUSIONS

The lack of any TL conceptual frameworks or reference models has facilitated the search and selection of results that align with traditionally accepted training principles and common sense, thus biasing our view and interpretation of the results. The bias in this field is unfortunately very evident. A recent review100 supported use of the ACWR (session rating of perceived exertion) but focused on only 2 of 13 identified studies. The authors failed to advise readers that the associations were in opposite directions, with Hulin et al1 reporting a higher injury risk at high ACWR (>1.5) and Jaspers et al90 showing a lower risk (ACWR >1.12). This kind of biased interpretation is regrettably quite common. The retrospective nature of the analysis and the several sources of researchers’ degrees of freedom facilitate false discoveries, P hacking, selective reporting, and HARKing. Unfortunately, based on the aforementioned methodologic concerns and the inability of studies to examine potential causal associations, practitioners should be cautious when deriving practical recommendations.

Although practitioners may be surprised and feel discouraged by our critical analysis, some good can come from our critique. Given the limitations presented in this 2-part series, we recommend that practitioners still rely on their clinical experience and intuition, combined with logical training principles and knowledge of physiological mechanisms and stimulus-inducing adaptations. Although theoretically under-training, “excessive” training, or both may be reasonably (but generally) considered predisposing factors to injury, this cannot be ascertained and quantified based on the literature. Training load-injury researchers should slow down and focus instead on producing higher-quality studies, even if this means many fewer studies (see Table for recommendations), including exploring more fundamental topics such as injury mechanisms in order to define frameworks that can be used to develop appropriate studies that are conducted according to established epidemiologic methods.

Note: During the preparation of this manuscript, new evidence and methodologic concerns have been presented.113,14,48 In light of these arguments, the Australian Institute of Sports has recently released a communication advising that the ACWR not be used as an indicator of injury risk (http://subscribe.ausport.gov.au/t-r-2CAD3E99D6C614D2540EF23F30FEDED).

### REFERENCES

1. American College of Sports Medicine. American College of Sports Medicine position stand. Progression models in resistance training

### Table. Minimum Recommendations Suggested for Future Authors Examining the Relationship of Training Load and Injury in Athletes

| Recommendations |
|------------------|
| Preregister prospective studies to limit P hacking. |
| Develop a hypothesis based on the conceptual frameworks and formulate a clear and defined research question. |
| Explicitly and honestly state the aim and nature of the prognostic study (exploratory, confirmatory, or explanatory or descriptive, predictive, or causal). |
| Define the training-load proxies based on the framework and justify their suitability for examining the research question. |
| Select the measure of exposures (eg, metrics) free of artifacts (valid). |
| Understand the number of observations and events needed for testing the hypothesis (eg, favoring multicenter studies) and limiting the risk of sparse data bias. |
| Adopt the appropriate statistical analysis, seeking advice in the planning phase of experts (eg, epidemiologists, statisticians). |
| Focus on the quality and consistency of reporting following international standards. |
| Discuss negative results (no selective reporting or selective discussion). |
| Interpret the results based on the real nature of the study (eg, descriptive versus explanatory) and the (reported) limitations. |
| Declare any conflict of interest. |
| Possibly provide (share) data, codes, or both. |
900
Volume 55  •  Number 9  •  September 2020
70. Bakhshi E, McArdle B, Mohammad K, et al. Let continuous outcome variables remain continuous. *Comput Math Methods Med*. 2012;2012:639124. doi:10.1155/2012/639124.
71. Bennette C, Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med Res Methodol*. 2012;12:21. doi:10.1186/1471-2288-12-21.
72. Caparros T, Casals M, Solana A, Pena J. Low external workload are related to higher injury risk in professional male basketball games. *J Sports Sci Med*. 2018;17(2):289–297.
73. McCall A, Jones M, Gelis L, et al. Monitoring loads and non-contact injury during the transition from club to National team prior to an international football tournament: a case study of the 2014 FIFA World Cup and 2015 Asia Cup. *J Med Sport*. 2018;21(8):800–804. doi:10.1016/j.jsams.2017.12.002.
74. Fanchini M, Rampinini E, Riggio M, Coats A, Pecci C, McCall A. Despite association, the acute : chronic work load ratio does not predict non-contact injury in elite footballers. *Sci Med Footh*. 2018;2(2):108–114. doi:10.1080/24733938.2018.1429014.
75. Bahr R, Clarsen B, Derman W, et al. International Olympic Committee consensus statement: methods for recording and reporting of epidemiological data on injury and illness in sport 2020 (including STROBE Extension for Sport Injury and Illness Surveillance (STROBE-SIIS)). *Br J Sports Med*. 2020;54(7):372–389. doi:10.1136/bjsports-2019-101969.
76. Hamilton GM, Meeuwisse WH, Emery CA, Shrier I. Examining the effect of the injury definition on risk factor analysis in circus artists. *Scand J Med Sci Sports*. 2012;22(3):330–334. doi:10.1111/j.1600-0838.2010.01245.x.
77. Sampson JA, Murray A, Williams S, et al. Injury risk-workload associations in NCAA American college football. *J Sci Med Sport*. 2018;21(12):1215–1220. doi:10.1016/j.jsams.2018.05.019.
78. Windt J, Ardern CL, Gabbert TJ, et al. Getting the most out of intensive longitudinal data: a methodological review of workload-injury studies. *BMJ Open*. 2018;8(10):e022626. doi:10.1136/bmjopen-2018-022626.
79. 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. doi:10.1136/bjsports-2018-099408.
80. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2012.
81. Murray NB, Gabbert TJ, Townshend AD. The use of relative speed zones in Australian football: are we really measuring what we think we are? *Int J Sports Physiol Perform*. 2018;13(4):442–451. doi:10.1123/ijssp.2017-0148.
82. Murray NB, Gabbert TJ, Townshend AD, Blanch P. Calculating acute : chronic workload ratios using exponentially weighted moving averages provide a more sensitive indicator of injury likelihood than rolling averages. *Br J Sports Med*. 2017;51(9):749–754. doi:10.1136/bjsports-2016-097152.
83. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol*. 2015;44(1):324–333. doi:10.1093/ije/dyu222.
84. Nielsen RO, Bertelsen ML, Ramskov D, et al. Time-to-event analysis for sports injury research part 2: time-varying outcomes. *Br J Sports Med*. 2019;53(1):70–78. doi:10.1136/bjsports-2018-100000.
85. Bahr R, Holme I. Risk factors for sports injuries—a methodological approach. *Br J Sports Med*. 2003;37(5):384–392. doi:10.1136/bjsm.37.5.384.
86. van Smeden M, de Groot JA, Moons KG, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med Res Methodol*. 2016;16(1):163. doi:10.1186/s12874-016-0267-3.
87. Riley RD, Snell KJ, Ensr J, et al. Minimum sample size for developing a multivariable prediction model: part II—binary and time-to-event outcomes [published correction appears in *Stat Med*. 2019;38(30):5672]. *Stat Med*. 2019;38(7):1276–1296. doi:10.1002/sim.7992.
88. Hoenig JM, Heisey DM. The abuse of power. *Am Stat*. 2001;55(1):19–24. doi:10.1198/000313001300339897.
89. Harel O, Mitchell EM, Perkins NJ, et al. Multiple imputation for incomplete data in epidemiologic studies. *Am J Epidemiol*. 2018;178(3):576–584. doi:10.1093/aje/kwx349.
90. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393. doi:10.1136/bmj.b2393.
91. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol*. 2017;9:157–166. doi:10.2147/CLEP.S129785.
92. Andridge RR, Little RJ. A review of hot deck imputation for survey non-response. *Int Stat Rev*. 2010;78(1):40–64. doi:10.1111/j.1751-5823.2010.00103.x.
93. Bennett DA. How can I deal with missing data in my study? *Aust N Z J Public Health*. 2001;25(5):464–469.
94. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869. doi:10.1136/bmj.c869.
95. Vandenbroucke JP, von Elm E, Altman DG, et al; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med*. 2007;147(8):W163–W194. doi:10.7326/0003-4819-147-8-200710160-00010-w1.
96. Nielsen RO, Bertelsen ML, Verhagen E, et al. When is a study result important for athletes, clinicians and team coaches/staff? *Br J Sports Med*. 2017;51(20):1454–1455. doi:10.1136/bjsports-2017-097759.
97. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: absolute risk reduction, relative risk reduction, and number needed to treat. *Perspect Clin Res*. 2016;7(1):51–53. doi:10.4103/2229-3485.173773.
98. Colby MJ, Dawson B, Peeling P, et al. Improvement of prediction and number needed to treat. *Ann Intern Med*. 2007;147(8):W163–W194. doi:10.7326/0003-4819-147-8-200710160-00010-w1.
99. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *J Clin Epidemiol*. 2015;68(2):134–143. doi:10.1016/j.jclinepi.2014.11.010.
100. Griffin A, Kenny IC, Comyns TM, Lyons M. The association between the acute : chronic workload ratio and injury and its application in team sports: a systematic review. *Sports Med*. 2019;50(3):561–580. doi:10.1007/s40279-019-01218-2.